TRANSACTIONS

OF THE

AMERICAN INSTITUTE

OF

ELECTRICAL ENGINEERS

VOLUME 77

1958

PART II. APPLICATIONS AND INDUSTRY

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1959 by the American Institute of Electrical Engineers
Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Preface

The AIEE *Transactions* for 1958 (volume 77) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics	Part II. Applications and Industry	Part III. Power Apparatus and Systems
Aircraft and Flight Test Instrumentation	Air Transportation	Insulated Conductors
Basic Sciences	Chemical Industry	Power Generation
Communication Switching Systems	Domestic and Commercial Applications	Power System Communications
Communication Theory	Electric Heating	Protective Devices
Computing Devices	Electric Welding	Relays
Data Communication	Feedback Control Systems	Rotating Machinery
Electrical Insulation	General Industry Applications	Substations
Electrical Techniques in Medicine and Biology	Industrial and Commercial Power Systems	Switchgear
Electronic and High Frequency Instruments	Industrial Control	System Engineering
Electronics	Industrial Power Rectifiers	Transformers
Indicating and Integrating Instruments	Land Transportation	Transmission and Distribution
Magnetic Amplifiers	Marine Transportation	
Nucleonic and Radiation Instrumentation	Metal Industry	
Nucleonics	Mining Industry	Education
Radio Communication Systems	Petroleum Industry	Research
Recording and Controlling Instruments	Production and Application of Light	Safety
Semiconductor Metallic Rectifiers		
Solid State Devices		
Special Instruments and Auxiliary Apparatus		
Telegraph Systems		
Telemetering		
Television and Aural Broadcasting Systems		
Wire Communication Systems		

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index. Volume 77 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., February 2–7, 1958.
2. Middle Eastern District Meeting, Washington, D. C., April 28–30, 1958.
3. Great Lakes District Meeting, East Lansing, Mich., May 5–7, 1958.
4. East Central District Meeting, Huntington, W. Va., May 13–15, 1958.
5. Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–26, 1958.
6. Pacific General Meeting, Sacramento, Calif., August 19–22, 1958.
7. Fall General Meeting, Pittsburgh, Pa., October 13–15, 1958.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Applications and Industry Index for 1958

1. Technical Subject Index

A-C Aircraft Generators, Computer Analysis of. Alger, Magnusson, Duane, Smith.....394–9
A-C Generating System, Analog Computer Simulation of an Aircraft Parallel. Riaz, Smith.....503–09
A-C Generators, A Static Exciter for Aircraft. Britten, Plette. (58–213).....271–6; disc. 276
A-C Generators, Coordination of Hydraulic Transmissions and. Aircraft. Caldwell, Coppinger, Smith.....377–86; disc. 386
A-C Tachometer, Approximate Analysis of the Drag Cup. Knox.....202–07; disc. 207
Advantages of a Pneumatic Electric Power System for High Mach Number Aircraft, The. Salvatti.....283–8
Air Circuit Breakers on a High-Capacity System, Field Testing of D-C. Hartsock, Scully, Stewart.....559–63; disc. 563
Air Filtration for Diesel Locomotives, Practical. Adams.....209–12
Air Intake, Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine. Kangas, Kern, Richardson.....230–5
(Air Transportation) Conductors and Conductor Protection. Steiner.....342–7; disc. 347
(Air Transportation) Effect of Operating Frequency on the Weight and Other Characteristics of Missile Alternators and Transformers. Turkington.....289–99; disc. 299
(Air Transportation) Equivalent Circuits for 3-phase 4-Wire Induction Motors Operating on Nonsymmetrical Systems. Mitchell.....468–75; disc. 475
(Air Transportation) External Heat Transfer from Cylindrical Self-Cooled Electric Equipment. Friedman.....372–7
(Air Transportation) Frequency Transient Measurements in Power Systems. Swanstrom.....312–16
(Air Transportation) Magnetic Circuit Properties of the Flux-Switch Inductor Alternator. Rauch.....307–12
(Air Transportation) Prediction of the Thermal Behavior of Blast-Cooled Generators by Use of a Digital Computer. Moroney.....388–94
(Air Transportation) Progress in the Thermal Design of Oil-Cooled Rotating Electric Machines. Richards.....330–4
(Air Transportation) The Turbonator: Design and Development. Duane, Harrison.....316–22
Air-Turbine Drives, An Electro-Hydraulic Speed and Load Division Control for Constant-Speed. Dantowitz, Norris.....99–106
Aircraft A-C Generators, A Static Exciter for. Britten, Plette. (58–213).....271–6; disc. 276
Aircraft A-C Generators, Coordination of Hydraulic Transmissions and. Caldwell, Coppinger, Smith.....377–86; disc. 386
Aircraft Alternators, Random Parallelling of. deMello, Rustebakke.....332–30; disc. 330
Aircraft Carriers, Special Power and Lighting Systems for. Cole, Garbaca.....241–5
Aircraft Electric Power Systems, Interrelationships Among Fault Torque, Voltage Unbalance, Fault Current, and Generator Impedances of. Jensen, Smith.....452–6; disc. 456
Aircraft Electrical Insulation, A Discussion of D-C High Potential Test Voltage for. Kilman, Dallas.....355–7; disc. 357
Aircraft Fluorescent Light System, Development of a Dimmable. Banios.....360–5
Aircraft Generators, Computer Analysis of A-C. Alger, Magnusson, Duane, Smith.....394–9
Aircraft, Lighting the Modern Commercial. Massie.....277–83
Aircraft Parallel A-C Generating System Analog Computer Simulation of an. Riaz, Smith.....503–09
Aircraft Power-Type Circuit Breaker for 600 F Ambient, Design of an. Austin.....338–42; disc. 342
Aircraft Relays, Design Considerations for. Massie, Kotnik.....305–07
Aircraft Switch and Relay Inductive-Circuit Test Loads. Dallas.....352–5
Aircraft Switching-Type Static Inverter Supplying Rotating Machines. Chirgwin.....334–8
Aircraft Systems, The Limitations of Induction Generators in Constant-Frequency. Edelyi, Kolatorowicz, Miller.....348–51; disc. 351
Aircraft, The Advantages of a Pneumatic Electric Power System for High Mach Number. Slavin.....283–8
Aircraft Use, A Brushless D-C Generator for. Duane.....365–71; disc. 371
Alternator, Magnetic Circuit Properties of the Flux-Switch Inductor. Rauch.....307–12
Alternators and Transformers, Effect of Operating Frequency on the Weight and Other Characteristics of Missile. Turkington.....289–99; disc. 299
Alternators, Random Paralleling of Aircraft. deMello, Rustebakke.....322–30; disc. 330
Amplitude- and Frequency-Sensitive Describing Function, Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an. Stein, Thaler.....91–5; disc. 96
Analog Computer Compensated Control Systems, The Design of. Bigelow.....409–15
Analog Computer Simulation of an Aircraft Parallel A-C Generating System. Riaz, Smith.....503–09
Analog Computer Study of a Torsional Vibrations Problem, An. Gray, McElhenny.....219–27
Analog Computer Study of Sampled-Data Systems. Chestnut, Thabal, Leiby.....634–40
Analysis and Stability Study of Finite Pulsed Feedback Systems, General. Farnanfarma.....148–62; disc. 162
Analysis and Synthesis of Feedback Control Systems, Graphical. Mitrovic:
Part I.....478–87
Part II.....487–96
Part III.....497–503; disc. 502
Analysis of A-C Aircraft Generators, Computer. Alger, Magnusson, Duane, Smith.....394–9
Analysis of a Guidance System, Time-Varying. Friedland.....75–81
Analysis of a Servomechanism with Backlash by the Ritz-Galerkin Method, An. Ogata, Atkinson.....82–4; disc. 84
Analysis of Cyclic-Rate Sampled-Data Feedback-Control Systems. Hufnagel.....421–3; disc. 423
Analysis of Ignitron Rectifiers for Reversing-Mill Drives. Hagensick, Cham.....259–68; disc. 268
Analysis of Multiple Sampler Systems with Finite Pulse Width, Open Loop. Farmanfarma.....20–8
Analysis of Sampled-Data Systems with Appreciable Pulse Width, Closed-Loop. Murphy, Kennedy.....659–64; disc. 664
Analysis of the Drag Cup A-C Tachometer, Approximate. Knox.....202–07; disc. 207
Analysis of Visual Elements in Roadway Lighting, An. Edman.....447–52
Analyzer Aids Design of Electrical Utility Automatic Dispatching System, Differential. Kirchmayer.....572–9; disc. 692
Application of Shift-Register Techniques to Materials Handling, The. Dieter.....624–8
Approximate Analysis of the Drag Cup A-C Tachometer. Knox.....202–07; disc. 207
Approximation of a Third-Order System, Synthesis of Control Systems Based on. Hausenbauer, Lago.....415–21; disc. 421
Arc-Back Current in Large Power Rectifiers, Mathematical Model and Procedure for. Fullerton, Teno.....456–64
Arc-Back Current in Large Power Rectifier Systems, Comparison of Calculated and Measured. Teno, Titus, Wagner.....585–9
Arm, Design of a Drive System for a Rotating. Fruchaa.....256–40; disc. 240
Artificial Cooling of Livestock in Hot Climates. Kelly, Bond, Ittner.....512–17
Automatic Control Systems, The Effect of Speed-Dependent Friction and Backlash on the Stability of. Freeman.....680–91; disc. 691
Automatic Dispatching System, Differential Analyzer Aids Design of Electrical Utility. Kirchmayer.....572–9; disc. 692
Automatic Hoist Controls, Trends in. Tiley, Zucker.....609–17; disc. 617
Automatic Speed-Control System for a Gravity Freight-Classification Yard, An. Bertl, Doch.....618–24
Automatic Warehousing, Static Control in. Boch, Fanthorp, Stuart.....595–602
Axles on The Chesapeake and Ohio Railway, Mobile Reflectoscopic Inspection of Railroad Car. Hauer, Angel.....143–7

B

Backlash by the Ritz-Galerkin Method, An Analysis of a Servomechanism with. Ogata, Atkinson.....82–4; disc. 84
Backlash Followed by a Dead Zone, The Describing Function of. Gronner.....403–08; disc. 408
Backlash on the Stability of Automatic Control Systems, The Effect of Speed-Dependent Friction and. Freeman.....680–91; disc. 691
Billet Heating, Classical Heat Flow Problems Applied to Induction. Baker.....106–12
Blast-Cooled Generators by Use of a Digital Computer, Prediction of the Thermal Behavior of. Moroney.....388–94
Braking, Quasi-Optimization of Relay Servomechanisms by Use of Stored Energy for. McDonald, Thaler.....628–34
Braking, Wheel-Slip Detection in Railroad. Hines.....215–19
Bridge Rectifiers with D-C Shunt Motor Load, Characteristics of Phase-Controlled. Pfaff.49–53; disc. 53
Brushless D-C Generator for Aircraft Use, A. Duane.....365–71; disc. 371

C

Cables Installed in the Same Residence: First Year of Comparative Operation, Heat Pump and Heating. Jones, Linden.....43–8
Cabooses, A Novel Generating System for Railroad. Haddad, Vercella, Brown.....566–70
Calculated and Measured Arc-Back Current in Large Power Rectifier Systems, Comparison of. Teno, Titus, Wagner.....585–9
Calculations for Power Rectifiers, Mathematical Model and Procedure for Arc-Back Current. Fullerton, Teno.....456–64
Carriers, Special Power and Lighting Systems on Aircraft. Cole, Garbaca.....241–5
Characteristic Equation of any Stable Linear Physical System, Two Theorems on the Number of Real Roots of the. Koenig.....670–3
Characteristics, Design, and Operation, Earth Source Heat Pumps: Freund, Whitlow.540–50; disc. 550
Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function, Obtaining the Frequency Response. Stein, Thaler.....91–5; disc. 96
Characteristics of an Electric Resistance Furnace Load. Stelzer.....570–2; disc. 691
Characteristics of Phase-Controlled Bridge Rectifiers with D-C Shunt Motor Load. Pfaff. 49–53; disc. 53
Chemical Process, System Considerations in Computer Control of a Semicontinuous. Stout.....640–51
Chesapeake and Ohio Railway, Mobile Reflectoscopic Inspection of Railroad Car Axles on The. Hauer, Angel.....143–7
Chopper Application, Silicon Transistor Performance in a. Giorgis, Thompson.....113–19; disc. 119
Circuit Breaker for 600 F Ambient, Design of an Aircraft Power-Type. Austin.....338–42; disc. 342
Circuit Breakers on a High-Capacity System, Field Testing of D-C Air. Hartsock, Scully, Stewart.....559–63; disc. 563
Circuit Properties of the Flux-Switch Inductor Alternator, Magnetic. Rauch.....307–12
Circuit Test Loads, Aircraft Switch and Relay Inductive-. Dallas.....352–5
Circuits, A Method of Scaling and Checking Computer. Lane.....67–70
Circuits for 3-Phase 4-Wire Induction Motor Operating in Nonsymmetrical Systems, Equivalent. Mitchell.....468–75; disc. 475
Classical Heat Flow Problems Applied to Induction Billet Heating. Baker.....106–12
Climates, Artificial Cooling of Livestock in Hot. Kelly, Bond, Ittner.....512–17
Closed-Loop Anlysis of Sampled-Data Systems with Appreciable Pulse Width. Murphy, Kennedy.....659–64; disc. 664
Closed-Loop Feedback Control Systems, Frequency Response of Nonlinear. Mikhail, Fett.....436–8
Closed-Loop Systems, The Sensitivity of the Poles of Linear. Huang.....182–6; disc. 186

Coils, Temperature of Conducting Ribbon Magnet. Douglas 465–8
Commercial Aircraft, Lighting the Modern. Massie 277–83
Comparison of Calculated and Measured Arc-Back Current in Large Power Rectifier Systems. Teno, Titus, Wagner 585–9
Compensating Saturation in Feedback Control Systems by Excess Error Storage. Chang, Archbald 16–20
Complex-Zero Signal Generator for Rapid System Testing. Lesdaris, Smith ..... 534–9; disc. 539
Component Variations, Limiting Values of Driving-Point Impedances and Transfer Functions Due to. Jones 38–40
Computer Analysis of A-C Aircraft Generators. Alger, Magnuson, Duane, Smith 394–9
Computer Circuits, A Method of Scaling and Checking. Lane 67–70
Computer Compensated Control Systems, The Design of Analog. Bigelow 409–15
Computer Control of a Semicontinuous Chemical Process, System Considerations in. Stout 640–51
Computer Control of Single-Loop and Multiloop Linear Systems, General Synthesis Procedure for. Kalman, Bertram 602–09
Computer Simulation of an Aircraft Parallel A-C Generating System Analog. Riar, Smith 503–09
Computer Study of a Torsional Vibrations Problem, An Analog. Gray, McElhenny 219–27
Computer Study of Sampled-Data Systems, Analog. Chestnut, Dabul, Leiby 634–40
Computer, Train Performance and Locomotive Tonnage Ratings Calculated by Digital. Hogan ..... 119–25; disc. 125
Conducting Ribbon Magnet Coils, Temperature of. Douglas 465–8
Conductors and Conductor Protection. Steiner ..... 342–7; disc. 347
Considerations for Aircraft Relays, Design. Massie, Kotnik 305–07
Considerations in Computer Control of a Semicontinuous Chemical Process, System. Stout 640–51
Constant-Frequency Aircraft Systems, The Limitations of Induction Generators in. Erdelyi, Kolatorowicz, Miller ..... 348–51; disc. 351
Constant-Speed Air-Turbine Drives, An Electro-Hydraulic Speed and Load Division Control for. Dantowitz, Norris 99–106
Constant-Speed Drive, Dynamic Representation of a Hydraulic. Smith 28–34
Contacts, Superposition Applied to Mechanical Rectifier. Dortort 167–72
Continuous Systems, Statistical Design Theory for Digital-Controlled. Chang ..... 191–8; disc. 198
Control Equipment on 100 New York City Subway Cars, First Year's Operating Experience with New. Weber ..... 34–7; disc. 37
Control for Constant-Speed Air-Turbine Drives, An Electro-Hydraulic Speed and Load Division. Dantowitz, Norris 99–106
Control, Motor-Operated Switch for Space Heating. Beckett 532–4
Control System, A Simple Nonlinearized. Zaborszky ..... 430–2; disc. 432
Control System Synthesis, Multivariable. Kavanagh ..... 425–9; disc. 429
Control Systems. Optimizing. Cosgriff, Emerling 13–16
Control Systems, Stability and Physical Realizability Considerations in the Synthesis of Multipole. Freeman 1–5
Control Systems, The Design of Analog Computer Compensated. Bigelow 409–15
Control Using Static-Frequency Doublers, Three-Phase Induction-Motor. Straughen, Biringer, Slemon ..... 58–66; disc. 66
Controls, Trends in Automatic Hoist. Tiley, Zucker ..... 609–17; disc. 617
Cooling of Livestock in Hot Climates, Artificial. Kelly, Bond, Ittner 512–17
Co-ordination of Hydraulic Transmissions and Aircraft A-C Generators. Caldwell, Coppinger, Smith ..... 377–86; disc. 386
Current Calculations for Power Rectifiers, Mathematical Model and Procedure for Arc-Back. Fullerton, Teno 456–64
Current in Large Power Rectifier Systems, Comparison of Calculated and Measured Arc-Back. Teno, Titus, Wagner 585–9
Cyclic-Rate Sampled-Data Feedback Control Systems, Analysis of. Hufnagel ..... 421–3; disc. 423
Cylindrical Self-Cooled Electric Equipment, External Heat Transfer from. Friedman 372–7

D

Data, Systems, Analog Computer Study of Sampled-. Chestnut, Dabul, Leiby 634–40
Data Systems with Appreciable Pulse Width, Closed-Loop Analysis of Sampled-. Murphy, Kennedy ..... 659–64; disc. 664
D-C Air Circuit Breakers on a High-Capacity System, Field Testing of. Harnack, Scully, Stewart ..... 559–63; disc. 563
D-C Generator for Aircraft Use, A Brushless. Duane ..... 365–71; disc. 371
D-C High Potential Test Voltage for Aircraft Electrical Insulation, A Discussion of. Kilman, Dallas ..... 335–7; disc. 357
D-C Reversing Mill Motor, Motor Field Control of Large. Hulls, Samuel ..... 579–85; disc. 692
D-C Shunt Motor Load, Characteristics of Phase Controlled Bridge Rectifiers with. Pfaff ..... 49–53; disc. 53
Dead Zone, The Describing Function of Backlash Followed by a. Gronner ..... 403–08; disc. 408
Derating Factor of Electric Motors that Drive Oil Well Pumping Units, Determining Actual Power Output and. Halderson ..... 551–7; disc. 557
Describing Function, Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive. Stein, Thaler ..... 91–5; disc. 96
Describing Function of Backlash Followed by a Dead Zone, The. Gronner ..... 403–08; disc. 408
Design and Development. The Turbonator: Duane, Harrison 316–22
Design Considerations for Aircraft Relays. Massie, Kotnik 305–07
Design Features of Large Semiconductor Rectifiers. Dortort ..... 249–53; disc. 253
Design of a Drive System for a Rotating Arm. Prucha ..... 236–40; disc. 240
Design of an Aircraft Power-Type Circuit Breaker for 600 V Ambient. Austin ..... 538–42; disc. 542
Design of Analog Computer Compensated Control Systems, The. Bigelow 409–15
Design Theory for Digital-Controlled Continuous Systems, Statistical. Chang ..... 191–8; disc. 198
Detection in Railroad Braking, Wheel-Slip. Hines 215–19
Determination of Output and Derating Factor of Electric Motors that Drive Oil Well Pumping Units. Halderson ..... 551–7; disc. 557
Development of a Dimmable Aircraft Flourescent Light System. Banios 360–5
Diesel-Electric Locomotives, An Evaluation of the Effect of Dirt on. Horine ..... 212–15; disc. 215
Diesel Locomotive Engine Air Intake, Field Experience with Electrostatic Precipitator on. Kangas, Kern, Richardson 230–5
Diesel Locomotives, Practical Air Filtration for. Adams 209–12
Differential Analyzer Aids Design of Electrical Utility Automatic Dispatching System. Kirchmayer ..... 572–9; disc. 692
Differential Reluctance Measuring System and Its Application to Model and Ship Testing, A. Wilson 245–9
Digital Computer, Prediction of the Thermal Behavior of Blast-Cooled Generators by Use of a. Moroney 388–94
Digital Computer, Train Performance and Locomotive Tonnage Ratings Calculated by. Hogan ..... 119–25; disc. 125
Digital-Controlled Continuous Systems, Statistical Design Theory for. Chang ..... 191–8; disc. 198
Dimmable Aircraft Fluorecent Light System, Development of a. Banios 360–5
Dirt on Diesel-Electric Locomotives, An Evaluation of the Effect of. Horine ..... 212–15; disc. 215
Discontinuous Feedback Control Systems, Optimum Response of. Neiline ..... 651–8; disc. 658
Discussion of D-C High-Potential Test Voltage for Aircraft Electrical Insulation, A. Kilman, Dallas ..... 355–7; disc. 357
Dispatching System, Differential Analyzer Aids Design of Electrical Utility Automatic. Kirchmayer ..... 572–9; disc. 692
Diversity of Farm Loads. Altman, Charity 439–43
Division Control for Constant-Speed Air-Turbine Drives, An Electro-Hydraulic Speed and Load. Dantowitz, Norris 99–106
Doppler Radar, Freight Car Tractive Resistance Measurements by. Campbell 563–6
Doublers, Three-Phase Induction-Motor Control Using Static-Frequency. Straughen, Biringer, Slemon ..... 58–66; disc. 66
Drag Cup A-C Tachometer, Approximate Analysis of the. Knox ..... 202–07; disc. 207
Drive, Dynamic Representation of a Hydraulic Constant-Speed. Smith 28–34
Drive System for a Rotating Arm, Design of a. Prucha ..... 236–40; disc. 240
Drives, An Electro-Hydraulic Speed and Load Division Control for Constant-Speed Air-Turbine, Dantowitz, Norris 99–106
Drives, Analysis of Ignitron Rectifiers for Reversing-Mill. Hagenah, Cham ..... 259–68; disc. 268
Driving-Point Impedances and Transfer Functions Due to Component Variations, Limiting Values of. Jones 38–40
Driving Source Impedance on Motor Performance, Effect of. Diamond ..... 529–31; disc. 531
Dual-Mode Servomechanism Utilizing Saturation Switching, A. Weed, Weimer ..... 590–3; disc. 593
Dynamic Representation of a Hydraulic Constant-Speed Drive. Smith 28–34

E

Earth Source Heat Pumps: Characteristics, Design and Operation. Freund, Whitlow ..... 540–50; disc. 550
Effect of Driving Source Impedance on Motor Performance. Diamond ..... 529–31; disc. 531
Effect of Operating Frequency on the Weight and Other Characteristics of Missile Alternators and Transformers. Turkington ..... 289–99; disc. 299
Effect of Quantization in Sampled-Feedback Systems. Bertram 177–82
Effect of Speed-Dependent Friction and Backlash on the Stability of Automatic Control Systems, The. Freeman ..... 680–91; disc. 691
Electric Equipment, External Heat Transfer from Cylindrical Self-Cooled. Friedman 372–7
(Electric Heating) Classical Heat Flow Problems Applied to Induction Billet Heating. Baker 106–12
Electric Machines, Progress in the Thermal Design of Oil-Cooled Rotating. Richards 330–4
Electric Power System for High Mach Number Aircraft, The Advantages of a Pneumatic. Slavin 283–8
Electric Power Systems, Interrelationships Among Fault Torque, Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft. Jensen, Smith ..... 452–6; disc. 456
Electric Resistance Furnace Load, Characteristics of an. Steizer ..... 570–2; disc. 691
Electric Systems, Metallic Rectifiers for Shipboard. Straub, Wiest 70–4
Electric Systems, Overcurrent Protection in Large Capacity Shipboard. Cole ..... 126–30; disc. 130
Electric Versus Gas or Oil, Residential Heating:. Smith ..... 524–8; disc. 528
Electrical Control of Gas Turbines in Gas Transmission Service, Yannone, Niemoller, Goertzen ..... 6–12; disc. 12
Electrical Insulation, A Discussion of D-C High Potential Test Voltage for Aircraft. Kilman, Dallas ..... 355–7; disc. 357
Electrical Insulation at Ultrahigh Temperatures, Properties of. Duncan, Felger ..... 300–05; disc. 305
Electrical Utility Automatic Dispatching System, Differential Analyzer Aids Design of. Kirchmayer ..... 572–9; disc. 692
Electricity for House Heating in the Tennessee Valley, The Use of. Martin 443–7
Electrification—An Approximate Method for Its Determination, The Financial Return on Railway. Brown, Marmaros ..... 163–5; disc. 165
(Electrification) Diversity of Farm Loads. Altman, Charity 439–43
Electro-Hydraulic Speed and Load Division Control for Constant-Speed Air-Turbine Drives, An. Dantowitz, Norris 99–106
Electrostatic Precipitator on Diesel Locomotive Engine Air Intake, Field Experience with. Kangas, Kern, Richardson 230–5
Elements in Roadway Lighting, An Analysis of Visual. Edman 44–52
Energy for Braking, Quasi-Optimization of Relay Servomechanisms by Use of Stored. McDonald, Thaler 628–34
Engine Air Intake, Field Experience with Electrostatic Precipitator on Diesel Locomotive. Kangas, Kern, Richardson 230–5
Equation of any Stable Linear Physical System, Two Theorems on the Number of Real Roots of the Characteristic. Koenig 670–3
Equipment on 100 New York City Subway Cars, First Year's Operating Experience with New Control. Weber ..... 34–7; disc. 37
Equipment, Ventilating Systems for Railway Traction. Widell 227–30
Equivalent Circuits for 3-Phase 4-Wire Induction Motors Operating on Nonsymmetrical Systems. Mitchell ..... 468–75; disc. 475
Error Storage, Compensating Saturation in Feedback Control Systems by Excess. Chang, Archbald 16–20
Evaluation of the Effect of Dirt on Diesel-Electric Locomotives, An. Horine ..... 212–15; disc. 215
Exciter for Aircraft A-C Generators, A Static. Britten, Piette. (58–213) ..... 271–6; disc. 276
External Heat Transfer from Cylindrical Self-Cooled Electric Equipment. Friedman 372–7

F

Factor of Electric Motors that Drive Oil Well Pumping Units, Determining Actual Power Output and Derating. Halderson ..... 551–7; disc. 557
Farm Loads, Diversity of. Altman, Charity 439–43

Fault Torque, Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft Electric Power Systems, Interrelationships Among. Jensen, Smith ..... 452-6; disc. 456
(Feedback Control) A Dual-Mode Servomechanism Utilizing Saturation Switching. Weed, Weimer. ..... 590-3; disc. 593
(Feedback Control) A Nonlinear Integrator for Servomechanisms. Clegg 41-2
(Feedback Control) A Simple Nonlinearized Control System. Zaborszky ..... 430-2; disc. 432
(Feedback Control) An Analysis of a Servomechanism with Backlash by the Ritz-Galerkin Method. Ogata, Atkinson ..... 82-4; disc. 84
(Feedback Control) Analog Computer Study of Sampled-Data Systems. Chestnut, Dabul, Leiby 634-40
(Feedback Control) Analysis of Multiple Sampler Systems with Finite Pulse Width, Open Loop. Farmanfarma 20-8
(Feedback Control) Approximate Analysis of the Drag Cup A-C Tachometer. Knox ..... 202-7; disc. 207
(Feedback Control) Closed-Loop Analysis of Sampled-Data Systems with Appreciable Pulse Width. Murphy, Kennedy ..... 659-64; disc. 664
(Feedback Control) Complex-Zero Signal Generator for Rapid System Testing. Lendaris, Smith ..... 534-9; disc. 539
(Feedback Control) Differential Analyzer Aids Design of Electrical Utility Automatic Dispatching System. Kirchmayer ..... 572-9; disc. 692
(Feedback Control) Dynamic Representation of a Hydraulic Constant-Speed Drive. Smith 28-34
(Feedback Control) Effect of Driving Source Impedance on Motor Performance. Diamond ..... 529-31; disc. 531
(Feedback Control) General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems. Kalman, Bertram 602-09
(Feedback Control) Instability in a Nonlinear Conditionally Stable System Subjected to a Sinusoidal Input. Douce, King 565-70
(Feedback Control) Limiting Values of Driving-Point Impedances and Transfer Functions Due to Component Variations. Jones 38-40
(Feedback Control) Multivariable Control System Synthesis. Kavanagh ..... 425-9; disc. 429
(Feedback Control) Optimizing Control Systems. Cosgriff, Emerling 13-16
(Feedback Control) Silicon Transistor Performance in a Chopper Application. Giorgis, Thompson ..... 113-19; disc. 119
(Feedback Control) Stability and Practical Reliability Considerations in the Synthesis of Multipole Control Systems. Freeman 1-5
Feedback Control, Stability Study of a Third-Order Servomechanism with Multiplicative. Ku, Chen 131-6
(Feedback Control) Statistical Design Theory for Digital-Controlled Continuous Systems. Chang ..... 191-8; disc. 198
(Feedback Control) Synthesis of Control Systems Based on Approximation to a Third-Order System. Hausenbauer, Lago ..... 415-21; disc. 421
(Feedback Control) System Considerations in Computer Control of a Semicontinuous Chemical Process. Stout 640-51
Feedback Control Systems, Analysis of Cyclic-Rate Sampled-Data. Hufnagel ..... 421-3; disc. 423
Feedback Control Systems by Excess Error Storage, Compensating Saturation in. Chang, Archbald 16-20
Feedback Control Systems, Frequency Response of Nonlinear Closed-Loop. Mikhail, Fett 436-8
Feedback Control Systems, Graphical Analysis and Synthesis of. Mitrovic:
Part I 478-87
Part II 487-96
Part III ..... 497-502; disc. 502
Feedback Control Systems, Optimum Response of Discontinuous. Nesline ..... 651-8; disc. 658
(Feedback Control) The Describing Function of Backlash Followed by a Dead Zone. Gronner ..... 403-08; disc. 408
(Feedback Control) The Design of Analog Computer Compensated Control Systems. Bigelow 409-15
(Feedback Control) The Effect of Speed-Dependent Friction and Backlash on the Stability of Automatic Control Systems. Freeman ..... 680-91; disc. 691
(Feedback Control) The Sensitivity of the Poles of Linear, Closed-Loop Systems. Huang ..... 182-6; disc. 186
(Feedback Control) Time-Varying Analysis of a Guidance System. Friedland 75-81
(Feedback Control) Transfer Function of 2-Phase Servomotors. Mikhail, Fett ..... 97-8; disc. 98
(Feedback Control) Two Theorems on the Number of Real Roots of the Characteristic Equation of any Stable Linear Physical System. Koenig 670-3
(Feedback Control) Use of Phase Space in Transient-Stability Studies. Bow, Van Ness ..... 187-91; disc. 191
Feedback Phase Space Trajectories and Application to Servo Systems, Positive. Meiksin 673-9
Feedback Systems, A Phase-Plane Approach to Relay Sampled-Data. Mullin, Jury ..... 517-21; disc. 521
Feedback Systems, General Analysis and Stability Study of Finite Pulsed. Farmanfarma 148-62; disc. 162
Feedback Systems, Staggered Sampling to Improve Stability of Multiple-Samper. Andeen ..... 399-402; disc. 402
Feedback Systems, The Effect of Quantization in Sampled-. Bertram 177-82
Field Control of Large D-C Reversing Mill Motors, Motor. Hulls, Samuel ..... 579-85; disc. 692
Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine Air Intake. Kangas, Kern, Richardson 230-5
Field Testing of D-C Air Circuit Breakers on a High-Capacity System. Harnsock, Scully, Stewart ..... 559-63; disc. 563
Filtration for Diesel Locomotives, Practical Air. Adams 209-12
Financial Return on Railway Electrification—An Approximate Method for Its Determination, The. Brown, Marmaros ..... 163-5; disc. 165
Finite Pulse Width, Open Loop, Analysis of Multiple Sampler Systems with. Farmanfarma 20-8
Finite Pulsed Feedback Systems, General Analysis and Stability Study of. Farmanfarma 148-62; disc. 162
First Years Operating Experience with New Control Equipment on 100 New York City Subway Cars. Weber ..... 34-7; disc. 37
Fluorescent Light System, Development of a Dimmable Aircraft. Banios 360-5
Flux-Switch Alternator, Magnetic Circuit Properties of the. Rauch 307-12
Forage Harvester, Tramp-Iron Remover for a Field. Jacob, Dobie 509-11
Four-Wire Induction Motors Operating on Nonsymmetrical Systems, Equivalent Circuits for 3-Phase. Mitchell ..... 468-75; disc. 475
Freight Car Tractive Resistance Measurements by Doppler Radar. Campbell 501-6
Freight-Classification Yard, An Automatic Speed-Control System for a Gravity. Berti, Douch 618-24
Frequency Doublers, Three-Phase Induction-Motor Control Using Static-. Straughen, Biringer, Siemon ..... 58-66; disc. 66
Frequency on the Weight and Other Characteristics of Missile Alternators and Transformers, Effect of Operating. Turkington ..... 289-99; disc. 299
Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function, Obtaining the. Stein, Thaler ..... 91-5; disc. 96
Frequency Response of Nonlinear Closed-Loop Feedback Control Systems. Mikhail, Fett 436-8
Frequency Transient Measurements In Power Systems. Swanstrom 312-16
Friction and Backlash on the Stability of Automatic Control Systems, The Effect of Speed-Dependent. Freeman ..... 680-91; disc. 691
Function of Time Produced by a Describing. The Describing. Gronner ..... 403-08; disc. 408
Function of 2-Phase Servomotors, Transfer. Mikhail, Fett ..... 97-8; disc. 98
Functions Due to Component Variations, Limiting Values of Driving-Point Impedances and Transfer. Jones 38-40
Furnace Load, Characteristics of an Electric Resistance. Steiner ..... 570-2; disc. 691

G

Gas or Oil, Residential Heating: Electric Versus. Smith ..... 524-8; disc. 528
Gas Turbines in Gas Transmission Service, Electrical Control of. Yannone, Niemoller, Goertsen ..... 6-12; disc. 12
General Analysis and Stability Study of Finite Pulsed Feedback Systems. Farmanfarma ..... 148-62; disc. 162
General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems. Kalman, Bertram 602-09
Generating System, Analog Computer Simulation of an Aircraft Parallel A-C. Rinz, Smith 503-09
Generating System for Railroad Cabooses, A Novel. Haddad, Vercelli, Brown 566-70
Generation, The Tie Between a Utility and an Industrial When the Industrial Has. Fawcett ..... 136-42; disc. 142
Generator for Aircraft Use, A Brushless D-C. Dunfield ..... 365-71; disc. 371
Generator for Rapid System Testing, Complex-Zero Signal. Lendaris, Smith ..... 534-9; disc. 539
Generator Impedances of Aircraft Electric Power Systems, Interrelationships Among Fault Torque, Voltage Unbalance, Fault Current, and. Jensen, Smith ..... 452-6; disc. 456
Generators, A Static Exciter for Aircraft A-C. Britten, Plette. (38-213) ..... 271-6; disc. 276
Generators by Use of a Digital Computer, Prediction of the Thermal Behavior of Blast-Cooled. Moroney 388-94
Generators, Computer Analysis of A-C Aircraft. Alger, Magnuson, Duane, Smith 394-9
Generators, Co-ordination of Hydraulic Transmissions and Aircraft A-C. Caldwell, Coppinger, Smith ..... 377-86; disc. 386
Generators in Constant-Frequency Aircraft Systems, The Limitations of Induction. Erdelyi, Kolatorowicz, Miller ..... 348-51; disc. 351
Graphical Analysis and Synthesis of Feedback Control Systems. Mitrovic:
Part I 476-87
Part II 487-96
Part III ..... 497-502; disc. 502
Gravity Freight-Classification Yard, An Automatic Speed-Control System for a. Berti, Douch 618-24
Guidance System, Time-Varying Analysis of a. Friedland 75-81

H

Harvester, Tramp-Iron Remover for a Field Forage. Jacob, Dobie 509-11
Heat Flow Problems Applied to Induction Billet Heating, Classical. Baker 106-12
Heat Pump and Heating Cables Installed in the Same Residence: Firm Year of Comparative Operation. Jones, Linden 43-8
Heat Pumps: Characteristics, Design and Operation, Earth Source. Freund, Whitlow ..... 540-50; disc. 550
Heat Transfer from Cylindrial Self-Cooled Electric Equipment, External. Friedman 372-7
Heating Control, Motor-Operated Switch for Space. Beckett 532-4
Heating: Electric Versus Gas or Oil, Residential. Smith ..... 524-8; disc. 528
Heating in the Tennessee Valley, The Use of Electricity for House. Martin 443-7
High-Capacity System, Field Testing of D-C Air Circuit Breakers on a. Harnsock, Scully, Stewart ..... 559-63; disc. 563
High Mach Number Aircraft, The Advantages of a Pneumatic Electric Power System for. Slavin 283-8
High-Potential Test Voltage for Aircraft Electrical Insulation, A Discussion of D-C. Kilnan, Dallas ..... 355-7; disc. 357
Hoist Controls, Trends in Automatic. Tilcy, Zucker ..... 609-17; disc. 617
Hot Climates, Artificial Cooling of Livestock in. Kelly, Bond, Ittner 512-17
House Heating in the Tennessee Valley, The Use of Electricity for. Martin 443-7
Hydraulic Constant-Speed Drive, Dynamic Representation of a. Smith 28-34
Hydraulic Speed and Load Division Control for Constant-Speed Air-Turbine Drives, An Electro-. Dantowitz, Norris 99-106
Hydraulic Transmissions and Aircraft A-C Generators, Co-ordination of. Caldwell, Coppinger, Smith ..... 377-86; disc. 386

I

Ignitron Rectifiers for Reversing-Mill Drives, Analysis of. Hagenick, Cham ..... 259-68; disc. 268
Impedance on Motor Performance, Effect of Driving Source. Diamond ..... 529-31; disc. 531
Impedances and Transfer Functions Due to Component Variations, Limiting Values of Driving-Point. Jones 38-40
Induction Billet Heating, Classical Heat Flow Problems Applied to. Baker 106-12
Induction Generators in Constant-Frequency Aircraft Systems, The Limitations of. Erdelyi, Kolatorowicz, Miller ..... 348-51; disc. 351
Induction-Motor Control Using Static-Frequency Doublers, Three-Phase. Straughen, Biringer, Siemon ..... 58-66; disc. 66
Induction Motors Operating on Nonsymmetrical Systems, Equivalent Circuits for 3-Phase 4-Wire. Mitchell ..... 468-75; disc. 475
Inductor Alternator, Magnetic Circuit Properties of the Flux-Switch. Rauch 307-12
(Industrial and Commercial Power Systems) Characteristics of an Electric Resistance Furance Load. Steiner ..... 570-2; disc. 691
(Industrial Control) A Method of Scaling and Checking Computer Circuits. Lance 767-70
(Industrial Control) Characteristics of Phase-Controlled Bridge Rectifiers with D-C Shunt Motor Load. Pfaff ..... 49-53; disc. 53
(Industrial Control) Static Control in Automatic Warehousing. Bosch, Fanthorp, Sutart 595-602
(Industrial Control) Temperature of Conducting Ribbon Magnet Coils. Douglas 465-8
(Industrial Control) The Application of Shift-Register Techniques to Materials Handling. Diener 624-8
(Industrial Power Rectifiers) Field Testing of D-C Air Circuit Breakers on a High-Capacity System. Harnsock, Scully, Stewart ..... 559-63; disc. 563

(Industrial Power Rectifiers) Superposition Applied to Mechanical Rectifier Contacts. Dortort....167–72
Industrial When the Industrial Has Generation, The Tie Between a Utility and an. Fawcett....136–42; disc. 142
Input, Instability in a Nonlinear Conditionally Stable System Subjected to a Sinusoidal. Douce, King....665–70
Inspection of Railroad Car Axles on the Chesapeake and Ohio Railway, Mobile Reflectoscopic. Hauer, Angel....143–7
Instability in a Nonlinear Conditionally Stable System Subjected to a Sinusoidal Input. Douce, King....665–70
Insulation, A Discussion of D-C High Potential Test Voltage for Aircraft Electrial. Kilman, Dallas....355–7; disc. 357
Insulation at Ultrahigh Temperatures, Properties of Electrical. Duncan, Felger....300–05; disc. 305
Integrator for Servomechanisms, A Nonlinear. Clegg....41–2
Interrelationships Among Fault Torque, Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft Electric Power Systems. Jensen, Smith....452–6; disc. 456
Inverter Supplying Rotating Machines, Aircraft Switching-Type Static. Chirgwin....334–8
Iron Remover for a Field Forage Harvester, Tramp-. Jacob, Dobie....509–11

L

(Land Transportation) A Novel Generating System for Railroad Cabooses. Haddad, Vercella, Brown....566–70
(Land Transportation) An Analog Computer Study of a Torsional Vibrations Problem. Gray, McElhenny....219–27
(Land Transportation) An Automatic Speed-Control System for a Gravity Freight-Classification Yard. Berti, Desch....618–24
(Land Transportation) An Evaluation of the Effect of Dirt on Diesel-Electric Locomotives. Horine....212–15; disc. 215
(Land Transportation) Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine Air Intake. Kangas, Kern, Richardson....230–5
(Land Transportation) First Year's Operating Experience with New Control Equipment on 100 New York City Subway Cars. Weber. 34–7; disc. 37
(Land Transportation) Freight Car Tractive Resistance Measurements by Doppler Radar. Campbell....563–6
(Land Transportation) Mobile Reflectoscopic Inspection of Railroad Car Axles on The Chesapeake and Ohio Railway. Hauer, Angel....143–7
(Land Transportation) Practical Air Filtration for Diesel Locomotives. Adams....209–12
(Land Transportation) Technical Aspects of Providing Service to Single-Phase 60-Cycle Railroad Loads. Nagel, Gabrielle....172–6
(Land Transportation) The Financial Return on Railway Electrification—An Approximate Method for Its Determination. Brown, Marmaros....163–5; disc. 165
(Land Transportation) Train Performance and Locomotive Tonnage Ratings Calculated by Digital Computer. Hogan....119–25; disc. 125
(Land Transportation) Ventilating Systems for Railway Traction Equipment. Widell....227–30
(Land Transportation) Wheel-Slip Detection in Railroad Braking. Hines....215–29
Light System, Development of a Dimmable Aircraft Fluorescent. Banice....360–5
Lighting, An Analysis of Visual Elements in Roadway. Edman....447–52
Lighting Systems on Aircraft Carriers, Special Power and. Cole, Garbacz....241–5
Lighting the Modern Commercial Aircraft. Masie....277–83
Limitations of Induction Generators in Constant-Frequency Aircraft Systems, The. Erdelyi, Kolatorowicz, Miller....348–51; disc. 351
Limiting Values of Driving-Point Impedances and Transfer Functions Due to Component Variations. Jones....38–40
Linear, Closed-Loop Systems, The Sensitivity of the Poles of. Huang....182–6; disc. 186
Linear Physical System, Two Theorems on the Number of Real Roots of the Characteristic Equation of Any Stable. Koenig....670–3
Linear Systems, General Synthesis Procedure for Computer Control of Single-Loop and Multiloop. Kalman, Bertram....602–09
Livestock in Hot Climates, Artificial Cooling of. Kelly, Bond, Ittner....512–17
Load, Characteristics of an Electric Resistance Furnace. Stelzer....570–2; disc. 691
Load Division Control for Constant-Speed Air-Turbine Drives, An Electro-Hydraulic Speed and. Dantowitz, Norris....99–106
Loads, Aircraft Switch and Relay Inductive-Circuit Test. Dallas....352–5
Loads, Diversity of Farm. Altman, Charity....439–43
Loads, Technical Aspects of Providing Service to Single-Phase 60-Cycle Railroad. Nagel, Gabrielle..172–6
Locomotive Engine Air Intake, Field Experience with Electrostatic Precipitator on Diesel. Kangas, Kern, Richardson....230–5
Locomotive Tonnage Ratings Calculated by Digital Computer, Train Performance and. Hogan....119–25; disc. 125
Locomotives, An Evaluation of the Effect of Dirt on Diesel-Electric. Horine....212–15; disc. 215
Locomotives, Practical Air Filtration for Diesel. Adams....209–12
Loop Analysis of Sampled-Data Systems with Appreciable Pulse Width, Closed-. Murphy, Kennedy....659–64; disc. 664

M

Mach Number Aircraft, The Advantages of a Pneumatic Electric Power System for High. Salvin....283–8
Magnet Coils, Temperature of Conducting Ribbon. Douglas....465–8
Magnetic Circuit Properties of the Flux-Switch Inductor Alternator. Rauch....307–12
(Marine Transportation) A Differential Reluctance Measuring System and Its Application to Model and Ship Testing. Wilson....245–9
(Marine Transportation) Special Power and Lighting Systems on Aircraft Carriers. Cole, Garbacz....241–5
Materials Handling, The Application of Shift-Register Techniques to. Diener....624–8
Mathematical Model and Procedure for Arc-Back Current Calculations for Power Rectifiers. Fullerton, Teno....456–64
Measurements by Doppler Radar, Freight Car Tractive Resistance. Campbell....563–6
Measurements in Power Systems, Frequency Transient. Swanstrom....312–16
Measuring System and Its Application to Model and Ship Testing, A Differential Reluctance. Wilson....245–9
Mechanical Rectifier Contacts, Superposition Applied to. Dortort....167–72
(Metal Industry) Motor Field Control of Large D-C Reversing Mill Motors. Hulls, Samuel....579–85; disc. 692
Metallic Rectifiers for Shipboard Electric Systems. Straub, Wiest....70–4
Method, An Analysis of a Servomechanism with Backlash by the Ritz-Galerkin. Ogata, Atkinson....82–4; disc. 84
Method of Scaling and Checking Computer Circuits, A. Lance....67–70
Mill Drives, Analysis of Ignitron Rectifiers for Reversing-. Hagenick, Cham....259–68; disc. 268
Mill Motors, Motor Field control of Large D-C Reversing. Hulls, Samuel....579–85; disc. 692
(Mining Industry) Trends in Automatic Hoist Controls. Tiley, Zucker....609–17; disc. 617
Missile Alternators and Transformers, Effect of Operating Frequency on the Weight and Other Characteristics of. Turkington....289–99; disc. 299
Mobile Reflectoscopic Inspection of Railroad Car Axles on the Chesapeake and Ohio Railway. Hauer, Angel....143–7
Model and Procedure for Arc-Back Current Calculations for Power Rectifiers, Mathematical. Fullerton, Teno....456–64
Motor Field Control of Large D-C Reversing Mill Motors. Hulls, Samuel....579–85; disc. 692
Motor Load, Characteristics of Phase-Controlled Bridge Rectifiers with D-C Shunt. Pfaff....49–53; disc. 53
Motor-Operated Switch for Space Heating Control. Beckett....532–4
Motor Performance, Effect of Driving Source Impedance on. Diamond....529–31; disc. 531
Motors that Drive Oil Well Pumping Units, Determining Actual Power Output and Derating Factor of Electric. Halderson....551–7; disc. 557.
Multiloop Linear Systems, General Synthesis Procedure for Computer Control of Single-Loop and. Kalman, Bertram....602–09
Multiple-Sampler Feedback Systems, Staggered Sampling to Improve Stability of. Andeen....399–402; disc. 402
Multiple Sampler Systems with Finite Pulse Width, Open Loop, Analysis of. Farmanfarma....20–8
Multiplicative Feedback Control, Stability Study of a Third-Order Servomechanism With. Ku, Chen....131–6
Multipole Control Systems, Stability and Physical Realizability Considerations in the Synthesis of. Freeman....1–5
Multivariable Control System Synthesis. Kavanagh....425–9; disc. 429

N

New York City Subway Cars, First Year's Operating Experience with New Control Equipment on 100. Weber....34–7; disc. 37
Nonlinear Closed-Loop Feedback Control Systems, Frequency Response of. Mikhail, Fett....436–8
Nonlinear Conditionally Stable System Subjected to a Sinusoidal Input, Instability in a. Douce, King....665–70
Nonlinear Integrator for Servomechanisms, A. Clegg....41–2
Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function, Obtaining the Frequency Response Characteristics of a. Stein, Thaler....91–5; disc. 96
Nonlinearized Control System, A Simple. Zaborszky....430–2; disc. 432
Nonsymmetrical Systems, Equivalent Circuits for 3-Phase 4-Wire Induction Motor and. Mitchell....468–75; disc. 475
Novel Generating System for Railroad Cabooses, A. Haddad, Vercella, Brown....566–70

O

Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function. Stein, Thaler....91–5; disc. 96
Oil-Cooled Rotating Electric Machines, Progress in the Thermal Design of. Richards....330–4
Oil, Residential Heating: Electric Versus Gas or. Smith....524–8
Oil Well Pumping Units, Determining Actual Power Output and Derating Factor of Electric Motors that Drive. Halderson....551–7; disc. 557
Open Loop, Analysis of Multiple Sampler Systems with Finite Pulse Width-. Farmanfarma....20–8
Operating Experience with New Control Equipment on 100 New York City Subway Cars, First Year's. Weber....34–7; disc. 37
Operation, Heat Pump and Heating Cables Installed in the Same Residence: First Year of Comparative. Jones, Linden....43–8
(Optimal Sampling System) General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems. Kalman, Bertram....602–09
Optimization of Relay Servomechanisms by Use of Stored Energy for Braking, Quasi-. McDonald, Thaler....628–34
Optimizing Control Systems. Chestnut, Eckerling....13–16
Optimum Response of Discontinuous Feedback Control Systems. Nealine....651–8; disc. 658
Overcurrent Protection in Large Capacity Shipboard Electric Systems. Cole....126–30; disc. 130

P

Paralleling of Aircraft Alternators, Random. deMello, Rustebakke....322–30; disc. 330
Performance in a Chopper Application, Silicon Transistor. Giorgis, Thompson....113–19; disc. 119
(Petroleum Industry) Determining Actual Power Output and Derating Factor of Electric Motors that Drive Oil Well Pumping Units. Halderson....551–7; disc. 557
(Petroleum Industry) Electrical Control of Gas Turbines in Gas Transmission Service. Yannone, Niemoller, Goertzen....6–12; disc. 12
(Petroleum Industry) Transient Protection Conditions in Pipe-Line Stations. Turner....85–90
Phase-Controlled Bridge Rectifiers with D-C Shunt Motor Load, Characteristics of. Pfaff....49–53; disc. 53
Phase-Plane Approach to Relay Sampled-Data Feedback Systems, A. Mullin, Jury....517–21; disc. 521
Phase Space in Transient-Stability Studies, Use of. Bow, Van Ness....187–91; disc. 191
Phase Space Trajectories and Application to Servo Systems, Positive Feedback. Meikain....673–9
Physical Realizability Considerations in the Synthesis of Multipole Control Systems, Stability and. Freeman....1–5
Physical System, Two Theorems on the Number of Real Roots of the Characteristic Equation of any Stable Linear. Koenig....670–3
Pipe-Line Stations, Transient Protection Conditions in. Turner....85–90
Pneumatic Electric Power System for High Mach Number Aircraft, The Advantages of a. Slavin....283–8

Poles of Linear, Closed-Loop Systems, The Sensitivity of the. Haung.....142–6; disc. 186
Positive Feedback Phase Space Trajectories and Application to Servo Systems. Meikan.....673–9
Power and Lighting Systems on Aircraft Carriers, Special. Cole, Garbacz.....241–5
Power Factor of Rectifiers. Schmidt.....53–7
Power Output and Derating Factor of Electric Motors that Drive Oil Well Pumping Units, Determining Actual. Halderson.....551–7; disc. 557
Power Rectifier Systems, Comparison of Calculated and Measured Arc-Back Current in Large. Teno, Tinu, Wagner.....585–9
Power Rectifiers, Mathematical Model and Procedure for Arc-Back Current Calculations for. Fullerton, Teno.....456–64
Power System for High Mach Number Aircraft, The Advantages of a Pneumatic Electric. Slavin.....283–8
Power Systems, Frequency Transient Measurements in. Swanstrom.....312–16
Power-Type Circuit Breaker for 600 F Ambient, Design of an Aircraft. Austin.....338–42; disc. 342
Practical Air Filtration for Diesel Locomotives. Adams.....309–12
Precipitator on Diesel Locomotive Engine Air Intake, Field Experience with Electrostatic. Kangas, Kern, Richardson.....230–5
Prediction of the Thermal Behavior of Blast-Cooled Generators by Use of a Digital Computer. Moroney.....388–94
Problems Applied to Induction Billet Heating, Classical Heat Flow. Baker.....106–12
Progress in the Thermal Design of Oil-Cooled Rotating Electric Machines. Richards.....330–4
Properties of Electrical Insulation at Ultrahigh Temperatures. Duncan, Felger.....300–05; disc. 305
Properties of the Flux-Switch Inductor Alternator, Magnetic Circuit. Rauch.....307–12
Protection Conditions in Pipe-Line Stations, Transient. Turner.....85–90
Protection, Conductors and Conductor. Steiner.....342–7; disc. 347
Protection in Large Capacity Shipboard Electric Systems, Overcurrent. Cole.....126–30; disc. 130
Pulse Width, Closed-Loop Analysis of Sampled-Data Systems with Appreciable. Murphy, Kennedy.....659–64; disc. 664
Pulse Width, Open Loop, Analysis of Multiple Sampler Systems with Finite. Farmanfarma.....20–8
Pulsed Feedback Systems, General Analysis and Stability Study of Finite. Farmanfarma.....148–62; disc. 162
Pump and Heating Cables Installed in the Same Residence: First Year of Comparative Operation, Heat. Jones, Linden.....43–8
Pumping Units, Determining Actual Power Output and Derating Factor of Electric Motors that Drive Oil Well. Halderson.....551–7; disc. 557
Pumps: Characteristics, Design, and Operation, Earth Source Heat. Freund, Whitlow.....540–50; disc. 550

Q

Quantization in Sampled-Feedback Systems, The Effect of. Bertram.....177–82
Quasi-Optimization of Relay Servomechanisms by Use of Stored Energy for Braking. McDonald, Thaler.....628–34

R

Radar, Freight Car Tractive Resistance Measurements by Doppler. Campbell.....563–6
Railroad Braking, Wheel-Slip Detection in. Hines.....215–19
Railroad Cabooses, A Novel Generating System for. Haddad, Vercella, Brown.....566–70
Railroad Car Axles on The Chesapeake and Ohio Railway, Mobile Reflectoscopic Inspection of. Haser, Angel.....143–7
Railroad Loads, Technical Aspects of Providing Service to Single-Phase 60-Cycle. Nagel, Gabrielle.....172–6
Railway Electrification—An Approximate Method for Its Determination, The Financial Return on. Brown, Marmaros.....163–5; disc. 165
Railway Traction Equipment, Ventilating Systems for. Widell.....227–30
Random Paralleling of Aircraft Alternators. deMello, Rustebakke.....322–30; disc. 330
Rapid System Testing, Complex-Zero Signal Generator for. Landaris, Smith.....534–9; disc. 539
Ratings Calculated by Digital Computer, Train Performance and Locomotive Tonnage. Hogan.....119–25; disc. 125
Realizability Considerations in the Synthesis of Multipole Control Systems, Stability and Physical. Freeman.....1–5
Rectifier Contacts, Superposition Applied to Mechanical. Dortort.....167–72
Rectifier Systems, Comparison of Calculated and Measured Arc-Back Current in Large Power. Teno, Tinu, Wagner.....585–9
Rectifiers, Design Features of Large Semiconductor. Dortort.....249–53; disc. 253
Rectifiers for Reversing-Mill Drives, Analysis of Ignitron. Hagenick, Cham.....259–68; disc. 268
Rectifiers, Mathematical Model and Procedure for Arc-Back Current Calculations for Power. Fullerton, Teno.....456–67
Rectifiers, Power Factor of. Schmidt.....53–7
Rectifiers with D-C Shunt Motor Load, Characteristic of Phase-Controlled Bridge. Pfaff.....49–53; disc. 53
Reflectoscopic Inspection of Railroad Car Axles on the Chesapeake and Ohio Railway, Mobile. Haser, Angel.....143–7
Register Techniques to Materials Handling, The Application of Shift-. Dieser.....624–8
Relay Inductive-Circuit Test Loads, Aircraft Switch and. Dallas.....352–5
Relay Sampled-Data Feedback Systems, A Phase-Plane Approach to. Mullin, Jury.....517–21; disc. 521
Relay Servomechanisms by Use of Stored Energy for Braking, Quasi-Optimization of. McDonald, Thaler.....628–34
Relays, Design Considerations for Aircraft. Maule, Kotnik.....305–07
Reluctance Measuring System and Its Application to Model and Ship Testing, A Differential. Wilson.....245–9
Remover for a Field Forage Harvester, Tramp-Iron. Jacob, Dobie.....509–11
Representation of a Hydraulic Constant-Speed Drive, Dynamic. Smith.....28–34
Residence: First Year of Comparative Operation, Heat Pump and Heating Cables Installed in the Same. Jones, Linden.....43–8
Residential Heating: Electric Versus Gas or Oil. Smith.....524–8; disc. 528
Resistance Furnace Load, Characteristics of an Electric. Stelzer.....570–2; disc. 691
Resistance Measurements by Doppler Radar, Freight Car Tractive. Campbell.....563–6
Response of Discontinuous Feedback Control Systems, Optimum. Nesline.....651–8; disc. 658
Response of Nonlinear Closed-Loop Feedback Control Systems, Frequency. Mikhail, Fett.....436–8
Reversing-Mill Drives, Analysis of Ignitron Rectifiers for. Hagenick, Cham.....259–68; disc. 268
Reversing Mill Motors, Motor Field Control of Large D-C. Hulls, Samuel.....579–85; disc. 692
Ribbon Magnet Coils, Temperature of Conducting. Douglas.....465–8
Ritz-Galerkin Method, An Analysis of a Servomechanism with Backlash by the. Ogata, Atkinson.....82–4; disc. 84
Roadway Lighting, An Analysis of Visual Elements in. Edman.....447–52
Roots of the Characteristic Equation of any Stable Linear Physical System, Two Theorems on the Number of Real. Koenig.....670–5
Rotating Arm, Design of a Drive System for a. Prucha.....236–41; disc. 240
Rotating Electrical Machines, Progress in the Thermal Design of Oil-Cooled. Richards.....330–4
Rotating Machines, Aircraft Switching-Type Static Inverter Supplying. Chirgwin.....334–8

S

Sampled-Data Feedback Control Systems, Analysis of Cyclic-Rate. Hufnagel.....421–3; disc. 423
Sampled-Data Feedback Systems, A Phase-Plane Approach to Relay. Mullin, Jury.....517–21; disc. 521
Sampled-Feedback Systems, The Effect of Quantization in. Bertram.....177–82
Sampled Data Systems, Analog Computer Study of. Chestnut, Rabul, Lelby.....634–40
Sampled-Data Systems with Appreciable Pulse Width, Closed-Loop Analysis of. Murphy, Kennedy.....659–64; disc. 664
Sampler Systems with Finite Pulse Width, Open Loop, Analysis of Multiple. Farmanfarma.....20–8
Sampling to Improve Stability of Multiple Sampler Feedback Systems, Staggered. Anderson.....399–402; disc. 402
Saturation in Feedback Control Systems by Excess Error Storage, Compensating. Chang, Archbald.....16–20
Saturation Switching, A Dual-Mode Servomechanism Utilizing. Weed, Weimer.....590–5; disc. 593
Scaling and Checking Computer Circuits, A Method of. Lane.....67–70
Self-Cooled Electric Equipment, External Heat Transfer from Cylindrical. Friedman.....372–7
Semiconductor Rectifiers, Design Features of Large. Dortort.....249–53; disc. 253
Sensitivity of the Poles of Linear, Closed-Loop Systems, The. Haung.....182–6; disc. 186
Service to Single-Phase 60-Cycle Railroad Loads, Technical Aspects of Providing. Nagel, Gabrielle.....172–6
Servo Systems, Positive Feedback Phase Space Trajectories and Application to. Meikin.....673–9
Servomechanism for an Amplitude- and Frequency-Sensitive Describing Function, Obtaining the Frequency Response Characteristics of a Nonlinear. Stein, Thaler.....91–5; disc. 96
Servomechanism Utilizing Saturation Switching, A Dual-Mode. Weed, Weimer.....590–3; disc. 593
Servomechanism with Backlash by the Ritz-Galerkin Method, An Analysis of a. Ogata, Atkinson.....82–4; disc. 84
Servomechanism with Multiplicative Feedback Control, Stability Study of a Third-Order. Ku, Chen.....131–6
Servomechanisms, A Nonlinear Integrator for. Clegg.....41–2
Servomechanisms by Use of Stored Energy for Braking, Quasi-Optimization of Relay. McDonald, Thaler.....628–34
Servomotors, Transfer Function of 2-Phase. Mikhail, Fett.....97–8; disc. 98
Shift-Register Techniques to Materials Handling, The Application of. Dieser.....624–8
Ship Testing, A Differential Measuring System and Its Application to Model and. Wilson.....245–9
Shipboard Electric Systems, Metallic Rectifiers for. Straub, Wiest.....70–4
Shipboard Electric Systems, Overcurrent Protection in Large Capacity. Cole.....126–30; disc. 130
Shunt Motor Load, Characteristics of Phase Controlled Bridge Rectifiers with D-C. Pfaff.....49–53; disc. 53
Signal Generator for Rapid System Testing, Complex-Zero. Landaris, Smith.....534–9; disc. 539
Silicon Transistor Performance in a Chopper Application. Giorgis, Thompson.....115–19; disc. 119
Simple Nonlinearized Control System, A. Zaborszky.....430–2; disc. 432
Simulation of an Aircraft Parallel A-C Generating System, Analog Computer. Rias, Smith.503–09
Single Loop and Multiloop Linear Systems, General Synthesis Procedure for Computer Control of. Kalman, Bertram.....602–09
Single-Phase 60-Cycle Railroad Loads, Technical Aspects of Providing Service to. Nagel, Gabrielle.....172–6
Sinusoidal Input, Instability in a Nonlinear Conditionally Stable System Subjected to a. Douce, King.....665–70
600 F Ambient, Design of an Aircraft Power-Type Circuit Breaker for. Austin.....338–42; disc. 342
60-Cycle Railroad Loads, Technical Aspects of Providing Service to Single-Phase. Nagel, Gabrielle.....172–6
Space Heating Control, Motor-Operated Switch for. Beckett.....538–4
Space in Transient-Stability Studies, Use of Phase. Bow, Van Ness.....187–91; disc. 191
Special Power and Lighting Systems on Aircraft Carriers. Cole, Garbacz.....241–5
Speed and Load Division Control for Constant-Speed Air-Turbine Drives, An Electro-Hydraulic. Dantowitz, Norris.....99–106
Speed-Control System for a Gravity Freight-Classification Yard, An Automatic. Berti, Doach.....618–24
Speed-Dependent Friction and Backlash on the Stability of Automatic Control Systems, The Effect of. Freeman.....680–91; disc. 691
Stability and Physical Realizability Considerations in the Synthesis of Multipole Control Systems. Freeman.....1–5
Stability Studies, Use of Phase Space in Transient-. Bow, Van Ness.....187–91; disc. 191
Stability Study of a Third-Order Servomechanism with Multiplicative Feedback Control. Ku, Chen.....131–6
Stability Study of Finite Pulsed Feedback Systems, General Analysis and. Farmanfarma.....148–62; disc. 162
Stable Linear Physical System, Two Theorems on the Number of Real Roots of the Characteristic Equation of any. Koenig.....670–5
Stable System Subjected to a Sinusoidal Input, Instability in a Nonlinear Conditionally. Douce, King.....665–70
Staggered Sampling to Improve Stability of Multiple-Sampler Feedback Systems. Andeen.....399–402; disc. 402
Static Control in Automatic Warehousing. Bosch, Fanthorp, Stuart.....595–602
Static Exciter for Aircraft A-C Generators, A. Britten, Piette.....271–6; disc. 276
Static-Frequency Doublers, Three-Phase Induction-Motor Control Using. Straughen, Biringer, Siemon.....58–66; disc. 66
Static Inverter Supplying Rotating Machines, Aircraft Switching-Type. Chirgwin.....334–8
Stations, Transient Protection Conditions in Pipe-Line. Turner.....85–90

Statistical Design Theory for Digital-Controlled Continuous Systems. Chang....191–8; disc. 198
Storage, Compensating Saturation in Feedback Control Systems by Excess Error. Chang, Archbald....16–20
Stored Energy for Braking, Quasi-Optimization of Relay Servomechanisms by Use of. McDonald, Thaler....628–34
Study of a Torsional Vibrations Problem, An Analog Computer. Gray, McElhenny....219–27
Study of Finite Pulsed Feedback Systems, General Analysis and Stability. Farmanfarma....148–62; disc. 162
Study of Sampled-Data Systems, Analog Computer. Chestnut, Dabul, Leiby....634–40
Subway Cars, First Year's Operating Experience with New Control Equipment on 100 New York City. Weber....34–7; disc. 37
Superposition Applied to Mechanical Rectifier Contacts. Dortort....167–72
Switch and Relay Inductive-Circuit Test Loads, Aircraft. Dallas....352–5
Switch for Space Heating Control, Motor Operated. Beckett....532–4
Switching, A Dual-Mode Servomechanism Utilizing Saturation. Weed, Weimer....590–3; disc. 593
Switching-Type Static Inverter Supplying Rotating Machines, Aircraft. Chirgwin....334–8
Synthesis, Multivariable Control System. Kavanagh....425–9; disc. 429
Synthesis of Control Systems Based on Approximation to a Third-Order System. Hausenbauer, Lago....415–21; disc. 421
Synthesis of Feedback Control Systems, Graphical Analysis and. Mitrovic:
Part I....478–87
Part II....487–96
Part III....497–502; disc. 502
Synthesis of Multipole Control Systems, Stability and Physical Realizability Considerations in the. Freeman....1–5
System and Its Application to Model and Ship Testing, A Differential Reluctance Measuring. Wilson....245–9
System Considerations in Computer Control of a Semicontinuous Chemical Process. Stout....640–51
System for a Gravity Freight-Classification Yard, An Automatic Speed-Control. Berti, Dosch....618–24
System for a Rotating Arm, Design of a Drive. Prucha....236–40; disc. 240
System Synthesis, Multivariable Control. Kavanagh....425–9; disc. 429
System, Synthesis of Control Systems Based on Approximation to a Third-Order. Hausenbauer, Lago....415–21; disc. 421
System Testing, Complex-Zero Signal Generator for Rapid. Lendaris, Smith....534–9; disc. 539
Systems, Optimizing Control. Cosgriff, Emerling....13–16
Systems with Finite Pulse Width, Open Loop, Analysis of Multiple Sampler. Farmanfarma....20–8

T

Tachometer, Approximate Analysis of the Drag Cup A-C. Knox....202–07; disc. 207
Technical Aspects of Providing Service to Single-Phase 60-Cycle Railroad Loads. Nagel, Gabrielle....172–6
Techniques to Materials Handling, The Application of Shift-Register. Diener....624–8
Temperature of Conducting Ribbon Magnet Coils. Douglas....465–8
Temperatures, Properties of Electrical Insulation at Ultrahigh. Duncan, Felger....300–05; disc. 305
Tennessee Valley, The Use of Electricity for House Heating in the. Martin....443–7
Test Loads, Aircraft Switch and Relay Inductive-Circuit. Dallas....352–5
Test Voltage for Aircraft Electrical Insulation, A Discussion of D-C High-Potential. Kilman, Dallas....355–7; disc. 357
Testing, Complex-Zero Signal Generator for Rapid System. Lendaris, Smith....534–9; disc. 539
Testing of D-C Air Circuit Breakers on a High-Capacity System, Field. Hartuock, Scully, Stewart....559–63; disc. 563
Theorems on the Number of Real Roots of the Characteristic Equation of Any Stable Linear Physical System, Two. Koenig....670–3
Theory for Digital-Controlled Continuous Systems, Statistical Design. Chang....191–8; disc. 198
Thermal Behavior of Blast-Cooled Generators by Use of a Digital Computer, Prediction of the. Moroney....388–94
Thermal Design of Oil-Cooled Rotating Electric Machines, Progress in the. Richards....330–4
Third-Order Servomechanism with Multiplicative Feedback Control, Stability Study of a. Ku, Chen....131–6
Third-Order System, Synthesis of Control Systems Based on Approximation to a. Hausenbauer, Lago....415–21; disc. 421
3-Phase 4-Wire Induction Motors Operating on Nonsymmetrical Systems, Equivalent Circuits for. Mitchell....468–75; disc. 475
Three-Phase Induction-Motor Control Using Static-Frequency Doublers. Straughen, Biringer, Siemon....58–66; disc. 66
Tie Between a Utility and an Industrial When the Industrial Has Generation, The. Fawcett....136–42; disc. 142
Time-Varying Analysis of a Guidance System. Friedland....75–81
Tonnage Ratings Calculated by Digital Computer, Train Performance and Locomotive. Hogan....119–25; disc. 125
Torque, Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft Electric Power Systems, Interrelationships Among Fault. Jensen, Smith....452–6; disc. 456
Torsional Vibrations Problem, An Analog Computer Study of a. Gray, McElhenny....219–27
Traction Equipment, Ventilating Systems for Railway. Widell....227–30
Tractive Resistance Measurements by Doppler Radar, Freight Car. Campbell....563–6
Train Performance and Locomotive Tonnage Ratings Calculated by Digital Computer. Hogan....119–25; disc. 125
Trajectories and Application to Servo Systems, Positive Feedback Phase Space. Melkin....673–9
Tramp-Iron Remover for a Field Forage Harvester. Jacob, Dobie....509–11
Transfer from Cyclindrical Self-Cooled Electric Equipment, External Heat. Friedman....372–7
Transfer Function of 2-Phase Servomotors. Mikhail, Fett....97–8; disc. 98
Transfer Functions Due to Component Variations, Limiting Values of Driving-Point Impedances and. Jones....38–40
Transformers, Effect of Operating Frequency on the Weight and Other Characteristics of Missile Alternators and. Turkington....289–99; disc. 299
Transient Measurements in Power Systems, Frequency. Swanstrom....312–16
Transient Protection Conditions in Pipe-Line Stations. Turner....85–90
Transient-Stability Studies, Use of Phase Space in. Bow, Van Ness....187–91; disc. 191
Transistor Performance in a Chopper Application, Silicon. Giorgis, Thompson....113–19; disc. 119
Transmission Service, Electrical Control of Gas Turbines in Gas. Yannone, Niemoller, Goertzen....6–12; disc. 12
Transmission and Aircraft A-C Generators, Co-ordination of Hydraulic. Caldwell, Coppinger, Smith....377–86; disc. 386
(Transportation, Air) Conductors and Conductor Protection. Steiner....342–7; disc. 347
(Transportation, Air) Effect of Operating Frequency on the Weight and Other Characteristics of Missile Alternators and Transformers. Turkington....289–99; disc. 299
(Transportation, Air) Equivalent Circuits for 3-Phase 4-Wire Induction Motors Operating on Nonsymmetrical Systems. Mitchell....468–75; disc. 475
(Transportation, Air) External Heat Transfer from Cylindrical Self-Cooled Electric Equipment. Friedman....372–7
(Transportation, Air) Frequency Transient Measurements in Power Systems. Swanstrom....312–16
(Transportation, Air) Magnetic Circuit Properties of the Flux-Switch Inductor Alternator. Rauch....307–12
(Transportation, Air) Prediction of the Thermal Behavior of Blast-Cooled Generators by Use of a Digital Computer. Moroney....388–94
(Transportation, Air) Progress in the Thermal Design of Oil-Cooled Rotating Electric Machines. Richards....330–4
(Transportation, Air) The Turbonator: Design and Development. Duane, Harrison....316–22
(Transportation, Land) A Novel Generating System for Railroad Cabooses. Haddad, Vercella, Brown....566–70
(Transportation, Land) An Analog Computer Study of a Torsional Vibrations Problem. Gray, McElhenny....219–27
(Transportation, Land) An Automatic Speed-Control System for a Gravity Freight Classification Yard. Berti, Dosch....618–24
(Transportation, Land) An Evaluation of the Effect of Dirt on Diesel-Electric Locomotives. Horine....212–15; disc. 215
(Transportation, Land) Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine Air Intake. Kangas, Kern, Richardson....230–5
(Transportation, Land) First Year's Operating Experience with New Control Equipment on 100 New York City Subway Cars. Weber....34–7; disc. 37
(Transportation, Land) Freight Car Tractive Resistance Measurements by Doppler Radar. Campbell....563–6
(Transportation, Land) Mobile Reflectoscopic Inspection of Railroad Car Axles on the Chesapeake and Ohio Railway. Hauer, Angel....143–7
(Transportation, Land) Practical Air Filtration for Diesel Locomotives. Adams....209–12
(Transportation, Land) Technical Aspects of Providing Service to Single-Phase 60-Cycle Railroad Loads. Nagel, Gabrielle....172–6
(Transportation, Land) The Financial Return on Railway Electrification—An Approximate Method for Its Determination. Brown, Marmaros....163–5; disc. 165
(Transportation, Land) Train Performance and Locomotive Tonnage Ratings Calculated by Digital Computer. Hogan....119–25; disc. 125
(Transportation, Land) Ventilating Systems for Railway Traction Equipment. Widell....227–30
(Transportation, Land) Wheel-Slip Detection in Railroad Braking. Hines....215–19
(Transportation, Marine) A Differential Reluctance Measuring System and Its Application to Mode and Ship Testing. Wilson....245–9
(Transportation, Marine) Design of a Drive System for a Rotating Arm. Prucha....236–40; disc. 240
(Transportation, Marine) Metallic Rectifiers for Shipboard Electric Systems. Straub, Wiest....70–4
(Transportation, Marine) Overcurrent Protection in Large Capacity Shipboard Electric Systems. Cole....126–30; disc. 130
(Transportation, Marine) Special Power and Lighting Systems on Aircraft Carriers. Cole, Garbacz....241–5
Trends in Automatic Heat Controls. Tilley, Zucker....699–17; disc. 617
Turbines in Gas Transmission Service, Electrical Control of Gas. Yannone, Niemoller, Goertzen....6–12; disc. 12
Turbonator: Design and Development, The. Duane, Harrison....316–22
2-Phase Servomotors, Transfer Function of. Mikhail, Fett....97–8; disc. 98
Two Theorems on the Number of Real Roots of the Characteristic Equation of any Stable Linear Physical System. Koenig....670–3

U

Use of Electricity for House Heating in the Tennessee Valley, The. Martin....443–7
Use of Phase Space in Transient-Stability Studies. Bow, Van Ness....187–91; disc. 191
Utility and an Industrial When the Industrial Has Generation, The Tie Between a. Fawcett....136–42; disc. 142
Utility Automatic Dispatching System, Differential Analyzer Aids Design of Electrical. Kirchmayer....572–9; disc. 692

V

Values of Driving-Point Impedances and Transfer Functions Due to Component Variations, Limiting. Jones....38–40
Ventilating Systems for Railway Traction Equipment. Widell....227–30
Vibrations Problem, An Analog Computer Study of a Torsional. Gray, McElhenny....219–27
Visual Elements in Roadway Lighting, An Analysis of. Edman....447–52
Voltage for Aircraft Electrical Insulation, A Discussion of D-C High Potential Test. Kilman, Dallas....355–7; disc. 357
Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft Electric Power Systems, Interrelationships Among Fault Torque. Jensen, Smith....452–6; disc. 459

W

Warehousing, Static Control in Automatic. Boach, Fanthrop, Stuart....595–602
Weight and Other Characteristics of Missile Alternators and Transformers, Effect of Operating Frequency on the. Turkington....289–99; disc. 299
Wheel-Slip Detection in Railroad Braking. Hines....215–19

Y

Yard, An Automatic Speed-Control System for a Gravity Freight-Classification. Berti, Dosch....618–24

Z

Zero Signal Generator for Rapid System Testing, Complex-. Lendaris, Smith....534–9; disc. 539

2. Author Index

Adams, M. B. Practical Air Filtration for Diesel Locomotives. (58–500) 209–12
Alger, J. R. M.; E. F. Magnuson, J. T. Duane, R. T. Smith. Computer Analysis of A-C Aircraft Generators. (58–908) 394–9
Allen, J. W. Disc. 357
Altman, Landy B., Jr.; L. F. Charity. Diversity of Farm Loads. (57–1159) 439–43
Andeen, Richard E. Staggered Sampling to Improve Stability of Multiple-Sampler Feedback Systems. (58–770) 399–402
Angel, C. M.; E. R. Hauer. Mobile Reflectoscopic Inspection of Railroad Car Axles on The Chesapeake and Ohio Railway. (58–238) 143–7
Archbald, R. W.; S. S. L. Chang. Compensating Saturation in Feedback Control Systems by Excess Error Storage. (58–92) 16–20
Anthanasades, Michael. Disc. 593
Atkinson, C. P.; K. Ogata. An Analysis of a Servomechanism with Backlash by the Ritz-Galerkin Method. (58–86) ..... 82–4; disc. 85
Austin, B. O. Design of an Aircraft Power-Type Circuit Breaker for 600 F Ambient. (58–832) ..... 338–42; disc. 342

B

Baker, R. M. Classical Heat Flow Problems Applied to Induction Billet Heating. (58–202) 106–12
Balke, R. L. Disc. 358
Banios, E. W. Development of a Dimmable Aircraft Fluorescent Light System. (58–839) 360–5
Beckett, J. C. Motor-Operated Switch for Space Heating Control. (58–1014) 532–4
Berti, R. J.; T. J. Dosch. An Automatic Speed-Control System for a Gravity Freight-Classification Yard. (58–1266) 618–24
Bertram, J. E. The Effect of Quantization in Sampled-Feedback Systems. (58–91) 177–82
Bertram, J. E.; R. E. Kalman. General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems. (58–1161) 602–09
Bertram, J. E.; R. E. Kalman. Disc. 162
Bigelow, Stephen C. The Design of Analog Computer Compensated Control Systems. (58–800) 409–15
Biringer, P. P.; A. Straughen, G. R. Slemon. Three-Phase Induction-Motor Control Using Static-Frequency Doublers. (56–68) ..... 55–66; disc. 66
Bond, T. E.; C. F. Kelly, N. R. Ittner. Artificial Cooling of Livestock in Hot Climates. (58–1007) 512–17
Bosch, L. L.; A. J. Fanthrop, J. W. Stuart. Static Control in Automatic Warehousing. (58–1285) 595–602
Bow, S. T.; J. E. Van Ness. Use of Phase Space in Transient-Stability Studies. (58–83) ..... 187–91; disc. 191
Brightman, F. P. Disc. 142
Britten, H. H.; D. L. Piette. A Static Exciter for Aircraft A-C Generators. (58–213) .271–6; disc. 276
Brown, D. W.; L. B. Haddad, R. A. Vercella. A Novel Generating System for Railroad Cabooses. (58–1159) 566–70
Brown, H. F.; H. E. Marmaros. The Financial Return on Railway Electrification—An Approximate Method for Its Determination. (58–32) ..... 163–5; disc. 165
Browning, S. C. Disc. 691

C

Caldwell, S. C.; T. E. Coppinger, R. T. Smith. Coordination of Hydraulic Transmissions and Aircraft A-C Generators. (58–848) ..... 377–86; disc. 387
Campbell, R. D. Freight Car Tractive Resistance Measurements by Doppler Radar. (58–1212) 563–6
Cham, J.; C. G. Hagenick. Analysis of Ignitron Rectifiers for Reversing-Mill Drives. (58–218) ..... 259–68; disc. 268
Chamulak, J. Disc. 254
Chang, S. S. L. Statistical Design Theory for Digital-Controlled Continuous Systems. (58–90) ..... 191–8; disc. 200
Chang, S. S. L. Disc. 656
Chang, S. S. L.; R. W. Archbald. Compensating Saturation in Feedback Control Systems by Excess Error Storage. (58–92) 16–20
Charity, Leon F.; L. B. Altman, Jr. Diversity of Farm Loads. (57–1159) 439–43
Chen, C. F.; Y. H. Ku. Stability Study of a Third-Order Servomechanism with Multiplicative Feedback Control. (57–1034) 131–6
Chestnut, Harold; A. Dabul, D. Leiby. Analog Computer Study of Sampled-Data Systems. (58–1324) 634–40
Chirgwin, K. M. Aircraft Switching-Type Static Inverter Supplying Rotating Machines. (58–828) 334–8
Chirgwin, K. M.; L. J. Stratton. Disc. 351
Clausing, C. I. Disc. 563
Clegg, J. C. A Nonlinear Integrator for Servomechanisms. (58–10) 41–2
Cole, J. R. Overcurrent Protection in Large-Capacity Shipboard Electric Systems. (58–132) ..... 126–30; disc. 131
Cole, J. R.; M. L. Garbacz. Special Power and Lighting Systems on Aircraft Carriers. (58–367) 241–5
Coppinger, T. E.; S. C. Caldwell, R. T. Smith. Coordination of Hydraulic Transmissions and Aircraft A-C Generators. (58–848) ..... 377–86; disc. 387
Cosgriff, R. L. Disc. 84
Cosgriff, R. L.; R. A. Emerling. Optimizing Control Systems. (57–778) 13–16
Curtis, L. B. Disc. 165

D

Dabul, Amadeo; H. Chestnut, D. Leiby. Analog Computer Study of Sampled-Data Systems. (58–1324) 634–40
Dallas, J. P. Aircraft Switch and Relay Inductive-Circuit Test Loads. (58–841) 352–5
Dallas, J. P.; L. B. Kilman. A Discussion of D-C High-Potential Test Voltage for Aircraft Electrical Insulation. (58–845) ..... 355–7; disc. 359
Dantowitz, P.; L. G. Norris. An Electro-Hydraulic Speed and Load Division Control for Constant-Speed Air-Turbine Drives. (58–138) 99–106
deMello, F. P.; H. M. Rustebakke. Random Paralleling of Aircraft Alternators. (58–830) ..... 322–30; disc. 330
DeRusso, Paul M.; F. W. Nesline, Jr. Disc. 198
Diamond, Albert. Effect of Driving Source Impedance on Motor Performance. (58–1028) 528–31
Diebold, E. J. Disc. 253
Diener, Henry C., Jr. The Application of Shift-Register Techniques to Materials Handling. (58–1166) 624–8
Dobie, J. B.; F. C. Jacob. Tramp-Iron Remover for a Field Forage Harvester. (58–1006) 509–11
Dortort, I. K. Design Features of Large Semiconductor Rectifiers. (58–516) ..... 249–53; disc. 256
Dortort, I. K. Superposition Applied to Mechanical Rectifier Contacts. (58–137) 167–72
Dosch, T. J.; R. J. Berti. An Automatic Speed-Control System for a Gravity Freight-Classification Yard. (58–1266) 618–24
Douce, J. L.; R. E. King. Instability in a Nonlinear Conditionally Stable System Subjected to a Sinusoidal Input. (58–1327) 665–70
Douglas, John F. H. Temperature of Conducting Ribbon Magnet Coils. (58–809) 465–8
Duane, J. T. A Brushless D-C Generator for Aircraft Use. (58–916) ..... 365–71; disc. 372
Duane, J. T.; J. R. M. Alger, E. F. Magnuson, R. T. Smith. Computer Analysis of A-C Aircraft Generators. (58–908) 394–9
Duane, J. T.; J. W. Harrison. The Turbonator: Design and Development. (58–791) 316–22
Duncan, J.; M. M. Felger. Properties of Electrical Insulation at Ultrahigh Temperatures. (58–115) ..... 300–05; disc. 305

E

Eastcott, P. deH. Disc. 617
Eckel, C. H. Disc. 558
Edman, W. H. An Analysis of Visual Elements in Roadway Lighting. (58–549) 447–52
Emerling, R. A.; R. L. Cosgriff. Optimizing Control Systems. (57–778) 13–16
Erdelyi, E.; R. E. Kolatorowicz, W. R. Miller. The Limitations of Induction Generators in Constant-Frequency Aircraft Systems. (58–831) ..... 348–51; disc. 351

F

Fanthrop, A. J.; L. L. Bosch, J. W. Stuart. Static Control in Automatic Warehousing. (58–1285) 595–602
Farmanfarma, G. Analysis of Multiple Sampler Systems with Finite Pulse Width, Open Loop. (57–1037) 20–8
Farmanfarma, G. General Analysis and Stability Study of Finite Pulsed Feedback Systems. (58–240) ..... 148–62; disc. 162
Farnsworth, G. B.; A. Schmidt, Jr., J. P. Smith. Disc. 254
Fawcett, D. V. The Tie Between a Utility and an Industrial When the Industrial Has Generation. (57–642) ..... 136–42; disc. 142
Felger, M. M.; G. I. Duncan. Properties of Electrical Insulation at Ultrahigh Temperatures. (58–115) ..... 300–05; disc. 305
Fett, G. H. Disc. 96
Fett, G. H.; S. L. Mikhail. Frequency Response of Nonlinear Closed-Loop Feedback Control Systems. (58–695) 436–8
Fett, G. H.; S. L. Mikhail. Transfer Function of 2-Phase Servomotors. (58–12) 97–8
Flugstad, Morris. Disc. 351
Ford, D. J.; T. W. See. Disc. 435
Freeman, E. A. The Effect of Speed-Dependent Friction and Backlash on the Stability of Automatic Control Systems. (58–1268) ..... 680–91; disc. 691
Freeman, Herbert. Stability and Physical Realizability Considerations in the Synthesis of Multipole Control Systems. (57–782) 1–5
Freund, E. A.; G. B. Whitlow. Earth Source Heat Pumps: Characteristics, Design, and Operation. (58–1010) ..... 540–50; disc. 551
Friedland, B. Time-Varying Analysis of a Guidance System. (58–82) 75–81
Friedman, Daniel. External Heat Transfer from Cylindrical Self-Cooled Electric Equipment. (58–857) 372–7
Fullerton, H. P.; J. Teno. Mathematical Model and Procedure for Arc-Back Current Calculations for Power Rectifiers. (58–226) 456–64

G

Gabrielle, A. F.; T. J. Nagel. Technical Aspects of Providing Service to Single-Phase 60-Cycle Railroad Loads. (57–1096) 172–6
Garbacz, M. L.; J. R. Cole. Special Power and Lighting Systems on Aircraft Carriers. (58–567) 241–5
Gerlitz, R. K. Disc. 37
Giorgis, J.; C. C. Thompson. Silicon Transistor Performance in a Chopper Application. (58–68) ..... 113–19; disc. 119
Godar, A. A.; W. J. Roa, Jr. Disc. 241
Goertzen, F. P.; R. A. Yannone, A. B. Niemoller. Electrical Control of Gas Turbines in Gas Transmission Service. (57–991) 6–12
Gray, R. T.; S. W. McElhenny. An Analog Computer Study of a Torsional Vibrations Problem. (58–502) 219–27
Gronner, Alfred D. The Describing Function of Backlash Followed by a Dead Zone. (58–771) ..... 403–08; disc. 409

H

Haddad, L. B.; R. A. Vercella, D. W. Brown. A Novel Generating System for Railroad Cabooses. (58–1159) 566–70
Haesler, W. K. Disc. 268
Hagenick, C. G.; E. J. Cham. Analysis of Ignitron Rectifiers for Reversing-Mill Drives. (58–218) ..... 259–68; disc. 268
Halderson, M. H. Determining Actual Power Output and Derating Factor of Electric Motors that Drive Oil Well Pumping Units. (58–465) ..... 551–7; disc. 558
Hanrahan, D. J. Disc. 299
Harrison, J. W.; J. T. Duane. The Turbonator: Design and Development. (58–791) 316–22
Hartsock, P. L.; T. J. Scully, V. N. Stewart. Field Testing of D-C Air Circuit Breakers on a High-Capacity System. (58–225) ..... 559–63; disc. 563
Hauer, E. R.; C. M. Angel. Mobile Reflectoscopic Inspection of Railroad Car Axles on The Chesapeake and Ohio Railway. (58–238) 143–7
Hausenbauer, C. R.; G. V. Lago. Synthesis of Control Systems Based on Approximation to a Third-Order System. (58–797) ..... 415–21; disc. 421
Heard, T. C. Disc. 12
Henshaw, M. D. Disc. 215
Hicken, T. L. Disc. 240
Hill, G. L. Disc. 358
Hines, C. M. Wheel-Slip Detection in Railroad Braking. (58–503) 215–19
Hogan, J. E. Train Performance and Locomotive Tonnage Ratings Calculated by Digital Computer. (58–190) ..... 119–25; disc. 126
Horine, J. W. An Evaluation of the Effect of Dirt on Diesel-Electric Locomotives. (58–504) 212–15
Howell, J. K. Disc. 557
Huang, R. Y. The Sensitivity of the Poles of Linear, Closed-Loop Systems. (58–60) ..... 182–6; disc. 187
Hufnagel, R. E. Analysis of Cyclic-Rate Sampled-Data Feedback-Control Systems. (58–803) ..... 421–3; disc. 423
Hulls, L. R.; G. H. Samuel. Motor Field Control of Large D-C Reversing Mill Motors. (58–1265) ..... 579–85; disc. 692

I

Ittner, N. R.; C. F. Kelly, T. E. Bond. Artificial Cooling of Livestock in Hot Climates. (58–1007) 512–17

Ivy, J. G. Disc. 617
Izawa, Keisuke. Disc. 521, 539

J

Jacob, F. C.; J. B. Dobie. Tramp-Iron Remover for a Field Forage Harvester. (58-1006) 509–11
Jensen, N. H.; R. T. Smith. Interrelationships Among Fault Torque, Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft Electric Power Systems. (58-918) ..... 452–6; disc. 456
Jones, C. W.; E. E. Linden. Heat Pump and Heating Cables Installed in the Same Residence: First Year of Comparative Operation. (58-64) 43–8
Jones, Sheldon. Limiting Values of Driving-Point Impedances and Transfer Functions Due to Component Variations. (58-81) 38–40
Jury, E. I. Disc. 402, 502
Jury, Eliahu I.; F. J. Mullin. A Phase-Plane Approach to Relay Sampled-Data Feedback Systems. (58-1002) ..... 517–21; disc. 521
Jury, E. I., F. J. Mullin. Disc. 423

K

Kalman, R. E. Disc. 199
Kalman, R. E.; J. E. Bertram. General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems. (58-1161) 602–09
Kalman, R. E.; J. E. Bertram. Disc. 162
Kargas, Pell; W. C. Kern, E. L. Richardson. Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine Air Intake. (58-505) 230–5
Kavanagh, R. J. Disc. 250–5
Kavanagh, R. J. Multivariable Control System Synthesis. (58-852) ..... 425–9; disc. 429
Kelly, C. F.; T. E. Bond, N. R. Ittner. Artificial Cooling of Livestock in Hot Climates. (58-1007) 512–17
Kennedy, Harold B.; G. J. Murphy. Closed-Loop Analysis of Sampled-Data Systems with Appreciable Pulse Width. (58-1303) ..... 659–64; disc. 665
Kern, W. C.; P. Kargas, E. L. Richardson. Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine Air Intake. (58-505) 230–5
Killin, A. M. Disc. 691
Kilman, L. B.; J. P. Dallas. A Discussion of D-C High-Potential Test Voltage for Aircraft Electrical Insulation. (58-845) ..... 355–7; disc. 359
King, R. E.; J. L. Douce. Instability in a Nonlinear Conditionally Stable System Subjected to a Sinusoidal Input. (58-1327) 665–70
Kinnen, E.; J. Tou. Disc. 424
Kirchmayer, L. K. Differential Analyzer Aids Design of Electrical Utility Automatic Dispatching System. (57-946) ..... 572–9; disc. 692
Klopfenstein, Arthur. Disc. 692
Knox, L. A. Approximate Analysis of the Drag Cup A-C Tachometer. (58-85) ..... 202–07; disc. 208
Koenig, J. F. Two Theorems on the Number of Real Roots of the Characteristic Equation of Any Stable Linear Physical System. (58-1267) 670–3
Kolatorowicz, E. E.; E. Erdelyi, W. R. Miller. The Limitations of Induction Generators in Constant-Frequency Aircraft Systems. (58-831) ..... 348–51; disc. 351
Kotnik, E. F.; L. E. Mansir. Design Considerations for Aircraft Relays. (57-503) 305–07
Ku, Y. H.; C. F. Chen. Stability Study of a Third-Order Servomechanism with Multiplicative Feedback Control. (57-1034) 131–6
Kurzweil, Fred, Jr. Disc. 522

L

Lago, G. V.; C. R. Hausenbauer. Synthesis of Control Systems Based on Approximation to a Third-Order System. (58-797) ..... 415–21; disc. 421
Lane, L. Jubin. A Method of Scaling and Checking Computer Circuits. (58-42) 67–70
Leiby, David; H. Chestnut, A Dabul. Analog Computer Study of Sampled-Data Systems. (58-1324) 634–40
Lendaris, G. G.; O. J. M. Smith. Complex-Zero Signal Generator for Rapid System Testing. (58-1027) ..... 534–9; disc. 539
Linden, E. E. Disc. 528
Linden, E. E.; C. W. Jones. Heat Pump and Heating Cables Installed in the Same Residence: First Year of Comparative Operation. (58-64) 43–8
Lynch, Robert D. Disc. 254

M

Magnusson, E. F.; J. R. M. Alger, J. T. Duane, R. T. Smith. Computer Analysis of A-C Aircraft Generators. (58-908) 394–9
Marmaros, H. E.; H. F. Brown. The Financial Return on Railway Electrification—An Approximate Method for Its Determination. (58-32) ..... 163–5; disc. 165
Martin, Buford H. The Use of Electricity for House Heating in the Tennessee Valley. (57-1161) 443–7
Massie, L. E.; E. F. Kotnik. Design Considerations for Aircraft Relays. (57-503) 305–07
Massie, Philip E. Lighting the Modern Commercial Aircraft. (57-458) 277–83
McDonald, C.; G. J. Thaler. Quasi-Optimization of Relay Servomechanisms by Use of Stored Energy for Braking. (58-1271) 628–34
McElhenny, S. W.; R. T. Gray. An Analog Computer Study of a Torsional Vibrations Problem. (58-502) 219–27
McGee, Richard R. Disc. 130
Meiksin, Z. H. Positive Feedback Phase-Space Trajectories and Application to Servo Systems. (58-1301) 673–9
Mikhail, S. L. Disc. 96
Mikhail, S. L.; G. H. Fett. Frequency Response of Nonlinear Closed-Loop Feedback Control Systems. (58-891) 436–8
Mikhail, S. L.; G. H. Fett. Transfer Function of 2-Phase Servomotors. (58-12) 97–8
Miller, W. R.; E. Erdelyi, E. E. Kolatorowics. The Limitations of Induction Generators in Constant-Frequency Aircraft Systems. (58-831) ..... 348–51; disc. 351
Mitchell, J. R. Equivalent Circuits for 3-Phase 4-Wire Induction Motors Operating on Nonsymmetrical Systems. (58-844) ..... 468–75; disc. 475
Mitrovic, Dusan. Graphical Analysis and Synthesis of Feedback Control Systems:
Part I. (58-860) 476–87
Part II. (58-861) 487–96
Part III. (58-862) ..... 497–502; disc. 502
Moroney, Richard M. Prediction of the Thermal Behavior of Blast-Cooled Conductors by Use of a Digital Computer. (58-919) 388–94
Mullin, Francis J.; E. I. Jury. A Phase-Plane Approach to Relay Sampled-Data Feedback Systems. (58-1002) ..... 517–21; disc. 521
Mullin, F. J.; E. I. Jury. Disc. 423
Murphy, Gordon J.; H. B. Kennedy. Closed-Loop Analysis with Appreciable Pulse Width. (58-1303) ..... 659–64; disc. 665

N

Nagel, T. J.; A. F. Gabrielle. Technical Aspects of Providing Service to Single-Phase 60-Cycle Railroad Loads. (57-1096) 172–6
Nesline, F. W., Jr. Optimum Response of Discontinuous Feedback Control Systems. (58-1192) ..... 651–8; disc. 658
Nesline, Frederick W., Jr. Disc. 191
Nesline, F. William, Jr. Disc. 429
Nesline, F. W., Jr.; F. M. DeRuso. Disc. 198
Niemoller, A. B.; R. A. Tannone, F. P. Goerzen. Electrical Control of Gas Turbines in Gas Transmission Service. (57-991) 6–12
Norris, L. G.; P. Dantowitz. An Electro-Hydraulic Speed and Load Division Control for Constant-Speed Air-Turbine Drives. (58-138) 99–106

O

Ogata, Katuhiko; C. P. Atkinson. An Analysis of a Servomechanism with Backlash by the Ritz-Galerkin Method. (58-86) ..... 82–4; disc. 85
Owen, T. B. Disc. 338

P

Padegs, Andris. Disc. 97
Penn, W. B. Disc. 358
Pfaff, R. W. Characteristics of Phase-Controlled Bridge Rectifiers with D-C Shunt Motor Load. (58-40) 49–53
Plette, D. L.; H. H. Britten. A Static Exciter for Aircraft A-C Generators. (58-213) ..... 271–6; disc. 276
Pohl, R. V. Disc. 268
Polster, N. E. Disc. 119
Prucha, H. E. Design of a Drive System for a Rotating Arm. (58-559) ..... 236–40; disc. 241

R

Rauch, S. E. Magnetic Circuit Properties of the Flux-Switch Inductor Alternator. (58-768) 307–12
Rias, M.; P. E. Smith, Jr. Analog Computer Simulation of an Aircraft Parallel A-C Generating System. (58-922) 503–09
Richards, P. B. Progress in the Thermal Design of Oil-Cooled Rotating Electric Machines. (58-790) 330–4
Richardson, E. L.; P. Kangas, W. C. Kern. Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine Air Intake. (58-505) 230–5
Roa, W. J., Jr.; A. A. Godar. Disc. 241
Robinson, A. S. Disc. 198
Rubinstein, H. A.; G. R. Slemon. Disc. 276
Rustebakke, H. M.; F. P. deMello. Random Paralleling of Aircraft Alternators. (58-830) ..... 322–30; disc. 330

S

Samuel, G. H.; L. R. Hulls. Motor Field Control of Large D-C Reversing Mill Motors. (58-1265) ..... 579–85; disc. 692
Saunders, R. M. Disc. 531
Savant, C. J., Jr. Disc. 421
Schmidt, A., Jr. Power Factor of Rectifiers. (58-220) 53–7
Schmidt, A., Jr.; G. B. Farnsworth, J. P. Smith. Disc. 254
Scott, Arnold H. Disc. 305
Scully, T. J.; P. L. Hartsock, V. N. Stewart. Field Testing of D-C Air Circuit Breakers on a High-Capacity System. (58-225) ..... 559–63; disc. 563
Sennhauser, G. J. Disc. 165
Slavin, Morton A. The Advantages of a Pneumatic Electric Power System for High Mach Number Aircraft. (58-910) 283–8
Slemon, G. R. Disc. 475
Slemon, G. R.; H. A. Rubinstein. Disc. 276
Slemon, G. R.; A. Straughen, P. P. Biringer. Three-Phase Induction-Motor Control Using Static-Frequency Doublers. (58-68) ..... 58–66; disc. 66
Smith, G. S. Residential Heating: Electric Versus Gas or Oil. (58-1012) ..... 524–8; disc. 529
Smith, George S. Disc. 550
Smith, J. P.; A. Schmidt, Jr., G. B. Farnsworth. Disc. 254
Smith, Otto J. M. Disc. 408, 434
Smith, O. J. M.; G. G. Lendaris. Complex-Zero Signal Generator for Rapid System Testing. (58-1027) ..... 534–9; disc. 539
Smith, P. E., Jr.; M. Rias. Analog Computer Simulation of An Aircraft Parallel A-C Generating System. (58-922) 503–09
Smith, Paul E., Jr. Dynamic Representation of a Hydraulic Constant-Speed Drive. (57-777) 28–34
Smith, R. E. Disc. 386
Smith, R. T.; J. R. M. Alger, E. F. Magnusson, J. T. Duane. Computer Analysis of A-C Aircraft Generators. (58-908) 394–9
Smith, R. T.; S. C. Caldwell, T. E. Coppinger. Coordination of Hydraulic Transmissions and Aircraft A-C Generators. (58-848) ..... 377–86; disc. 387
Smith, R. T.; N. H. Jensen. Interrelationships Among Fault Torque, Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft Electric Power Systems. (58-918) ..... 452–6; disc. 456
Snively, H. N. Disc. 692
Stair, J., Jr. Disc. 125, 165
Stein, W. A.; G. J. Thaler. Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function. (58-84) ..... 91–5; disc. 96
Steiner, R. Conductors and Conductor Protection. (58-836) ..... 342–7; disc. 347
Steiner, W. A. Characteristics of an Electric Resistance Furnace Load. (58-1258) ..... 570–2; disc. 692
Stewart, V. N.; P. L. Hartsock, T. J. Scully. Field Testing of D-C Air Circuit Breakers on a High-Capacity System. (58-225) ..... 559–63; disc. 563
Stineman, R. W. Disc. 276, 347
Stout, T. M. System Considerations in Computer Control of a Semicontinuous Chemical Process. (58-1321) 640–51
Stratton, L. J.; K. M. Chirgwin. Disc. 351
Straub, C. L.; H. G. Wiest. Metallic Rectifiers for Shipboard Electric Systems. (58-135) 70–4
Straughen, A.; P. P. Biringer, G. R. Slemon. Three-Phase Induction-Motor Control Using Static-Frequency Doublers. (58-68) ..... 58–66; disc. 66
Stuart, J. W.; L. L. Bosch, A. J. Fanthrop. Static Control in Automatic Warehousing. (58-1285) 595–602
Swanstrom, G. N. Frequency Transient Measurements in Power Systems. (58-788) 312–16
See, T. W.; D. J. Ford. Disc. 433

T

Tauber, H. Disc. 691
Teno, J.; H. P. Fullerton. Mathematical Model and Procedure for Arc-Back Current Calculations for Power Rectifiers. (58-226) 456–64
Teno, J.; C. H. Titus, R. N. Wagner. Comparison of Calculated and Measured Arc-Back Current in Large Power Rectifier Systems. (58-223) 585–9

Thaler, G. J.; C. McDonald. Quasi-Optimization of Relay Servomechanisms by Use of Stored Energy for Braking. (58–1271) 628–34
Thaler, G. J.; W. A. Stein. Obtaining the Frequency-Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function. (58–84) ..... 91–5; disc. 96
Thompson, C. C.; J. Giorgis. Silicon Transistor Performance in a Chopper Application. (58–88) ..... 115–19; disc. 119
Tiley, G. L.; E. Zucker. Trends in Automatic Hoist Controls. (58–1201) ..... 609–17; disc. 617
Titus, C. H.; J. Teno, R. N. Wagner. Comparison of Calculated and Measured Arc-Back Current in Large Power Rectifier Systems. (58–223) ... 585–9
Tou, Julius. Disc. 403, 664
Tou, J.; E. Kinnen. Disc. 424
Tracht, J. W. Disc. 255
Turkington, R. E. Effect of Operating Frequency on the Weight and Other Characteristics of Missile Alternators and Transformers. (58–772) ..... 289–99; disc. 299
Turner, E. B. Transient Protection Conditions in Pipe-Line Stations. (57–990) 85–90

U

Ur, Hanoch. Disc. 186

V

Van Ness, J. E.; S. T. Bow. Use of Phase Space in Transient-Stability Studies. (58–83) ..... 187–91; disc. 191
Vercella, R. A.; L. B. Haddad, D. W. Brown. A Novel Generating System for Railroad Cabooses. (58–1159) 566–70

W

Wagner, R. N.; J. Teno, C. H. Titus. Comparison of Calculated and Measured Arc-Back Current in Large Power Rectifier Systems. (58–223) ... 585–9
Walter, R. K. Disc. 371
Wasmund, J. A. Disc. 241
Weber, G. W. First Year's Operating Experience with New Control Equipment on 100 New York City Subway Cars. (58–250) ..... 34–7; disc. 37
Weed, H. R.; F. C. Weimer. A Dual-Mode Servomechanism Utilizing Saturation Switching. (58–1191) ..... 590–3; disc. 595
Weimer, F. C.; H. R. Weed. A Dual-Mode Servomechanism Utilizing Saturation Switching. (58–1191) ..... 590–3; disc. 595
Weis, Gerald. Disc. 207
Whitlow, G. S.; E. A. Freund. Earth Source Heat Pumps: Characteristics, Design, and Operation. (58–1010) ..... 540–50; disc. 551
Widell, B. A. Ventilating Systems for Railway Traction Equipment. (58–507) 227–30
Wiest, H. G.; C. L. Stroub. Metallic Rectifiers for Shipboard Electric Systems. (58–135) 70–4
Wilson, B. J. Disc. 456
Wilson, J. T. R. Disc. 53, 66
Wilson, Meredith W. A Differential Reluctance Measuring System and Its Application to Model and Ship Testing. (58–548) 245–9
Winograd, H. Disc. 255
Wood, A. J. Disc. 330

Y

Yannone, R. A.; A. B. Niemoller, F. P. Goertsen. Electrical Control of Gas Turbines in Gas Transmission Service. (57–991) 6–12

Z

Zaborszky, J. A Simple Nonlinearized Control System. (58–894) ..... 430–2; disc. 435
Zucker, E.; G. L. Tiley. Trends in Automatic Hoist Controls. (58–1201) ..... 609–17; disc. 617

Stability and Physical Realizability Considerations in the Synthesis of Multipole Control Systems

HERBERT FREEMAN
MEMBER AIEE

Synopsis: With the use of the canonical form for a multipole control system as the starting point, the characteristic equation of the system is derived and its roots related to the poles and zeros of the given plant. Conditions are established which must be imposed on the plant to permit the design of a stable system having an arbitrary response. This is followed by the development of a technique which, by imposing constraints on the over-all response, enables the system designer to achieve a stable design even when certain of the plant's poles and zeros lie in an unstable region of the complex frequency plane. The method permits the direct stability design of multipole control systems in a manner similar to that presently possible for the conventional 2-pole systems. It is applicable to continuous-data systems as well as to sampled-data systems.

The conditions for the physical realizability of the controllers of a multipole control system designed for a specified over-all response are investigated, and it is shown that, except for the introduction of certain delays, physical realizability can always be achieved.

MULTIPOLE control systems are control systems having a multiplicity of independent inputs and dependent outputs. Since such systems involve a very large number of variables, their stability design is usually a very formidable problem. In the past, the approach taken to their synthesis has been to resort to trial-and-error studies on analog computers. Although the problem of multipole control system synthesis has received increased attention in recent years,[1-2] relatively little has been accomplished with respect to extending the accepted techniques of 2-pole system synthesis to the more general multipole case. Golomb and Usdin[3] considered the stability of a restricted class of multipole systems from the analysis point of view. Povejsil and Fuchs[4] developed a synthesis method which relied heavily on trial-and-error techniques and the aid of an analog computer. A synthesis procedure based on the adoption of certain variable-reducing system constraints (including a canonical form) is described by the author in a previous paper.[4] The development of the stability equations and the conditions for physical realizability on the basis of this synthesis procedure form the subject of this paper.

In recent years digital computers have been used increasingly as control elements in complex control systems. This is particularly true for the more complex multipole control systems for which the method described here is of particular importance. For this reason a certain emphasis is placed on systems of the sampled-data type. The method is, however, completely general, and applies equally to continuous-data systems and sampled-data systems.

Multipole Control Systems

A generalized block diagram of a linear time-invariant multipole control system having m independent inputs and n controlled outputs is shown in Fig. 1. The system is seen to consist of plant P (the physical equipment to be controlled) which is assumed to be unalterable, the forward controller D, the feedback controller H, and an input-output error-taking device. The symbols R, E, X, Y, and C denote column matrices, each representing a multiplicity of data quantities of the multipole control system. The lower-case letters associated with these symbols indicate the number of data quantities represented by each of the data lines. The plant matrix P is an $m \times n$ matrix whose elements are the various plant input-output transfer functions. Similarly, D and H are transfer matrices for the forward and feedback controllers, respectively. In the synthesis problem, D and H must be designed on the basis of a given plant and a specified over-all response.

The $m \times n$ structure of the plant matrix P suggested the desirability of defining a set of m-n "virtual" outputs and partitioning the system matrices into a and b channels, according to whether they are associated with the n real outputs or the m-n virtual ones. The resulting system, shown in Fig. 2, may be defined as a canonical form of multipole control system. The generalized plant matrix G of Fig. 2 is related by definition to the plant matrix P as follows

$$G = \begin{Vmatrix} P_a & P_c \\ 0 & I \end{Vmatrix} \tag{1}$$

where P_a is an $n \times n$ square matrix and I is the identity matrix.

On the basis of the canonical form, which in effect establishes a framework for an analytical description of a multipole control system, a synthesis procedure can be developed. The number of variables can be reduced without unduly restricting the freedom of the system designer by: 1. ruling out coupling between primary and secondary data channels in the feedback controller, and 2. by requiring independence of output restoration.* As a result of the first, the terms H_c and H_d of the matrix

$$H = \begin{Vmatrix} H_a & H_c \\ H_d & H_b \end{Vmatrix} \tag{2}$$

go to zero, and as a result of the second

$$P_a D_a H_a = \text{diagonal} \tag{3}$$

$$P_b D_b = \text{diagonal} \tag{4}$$

$$D_d = 0 \tag{5}$$

$$P_a D_c + P_c D_b = 0 \tag{6}$$

It now becomes feasible to solve the system equations explicitly for the two system output matrices (real and virtual). By defining an over-all system transfer matrix K according to the equation

$$C = KR \tag{7}$$

where

$$K = \begin{Vmatrix} K_a & K_c \\ K_d & K_b \end{Vmatrix} \tag{8}$$

and where

Paper 57-782, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24-28, 1957. Manuscript submitted March 19, 1957; made available for printing June 11, 1957.

HERBERT FREEMAN is with Sperry Gyroscope Company, Division of Sperry Rand Corporation, Great Neck, N. Y.

This work was part of a doctoral research program carried out by the author while on leave at Columbia University, New York, N. Y. Appreciation is expressed to Professor J. R. Ragazzini for his guidance. Some of the early parts of this research were supported by the U. S. Air Force under contract no. *AF 18(600)-677*, monitored by the Office of Scientific Research, Air Research and Development Command.

* This is the requirement that the solution of the homogeneous equations for the outputs, with an initial condition on one output, be zero except for the output subjected to the initial condition. For a more detailed description refer to reference 4.

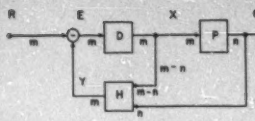
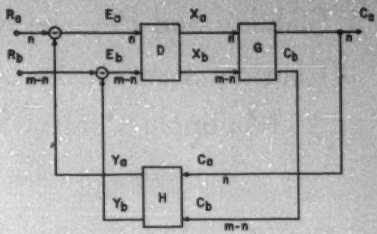


Fig. 1 (left). General system block diagram

Fig. 2 (right). Canonical form of multipole control system



$$C_a = K_a R_a + K_c R_b \tag{9}$$

$$C_b = K_d R_a + K_b R_b \tag{10}$$

one obtains

$$K_a = (I + P_a D_a H_a)^{-1} P_a D_a \tag{11}$$

$$K_b = (I + D_b H_b)^{-1} D_b \tag{12}$$

$$K_c = (I + P_a D_a H_a)^{-1} (P_a D_c + P_c D_b) \times H_b (I + D_b H_b)^{-1} H_b^{-1} \tag{13}$$

$$K_d = 0 \tag{14}$$

The Characteristic Equation

According to equation 7, a multipole system can be characterized by an over-all transfer matrix K whose elements are the transfer functions which relate the various outputs and inputs. It follows that for a system to be stable, all elements of its matrix K must be individually stable; i.e., their poles must lie in the left-half plane of the complex variable s, or in the interior of the unit circle of the z-plane ($z = e^{sT}$) in the case of sampled-data systems.[4-7]

The stability of a system as a whole can be established from the roots of the system characteristic equation. Consider first the case where $m = n$. From equations 8 and 11,

$$K = K_a = (I + P_a D_a H_a)^{-1} P_a D_a \tag{15}$$

$$= [I - (I + P_a D_a H_a)^{-1}] H_a^{-1} \tag{16}$$

$$(K_a)_{ij} = \sum_{k=1}^{n} [I - (I + P_a D_a H_a)^{-1}]_{ik} \frac{|H_{jk}|}{|H_a|} \tag{17}$$

$$= \sum_{k=1}^{n} \frac{[\delta_{ik} |I + P_a D_a H_a| - |(I + P_a D_a H_a)_{ki}|] |H_{jk}|}{|I + P_a D_a H_a| \, |H_a|} \tag{18}$$

Since every element of the cofactors $|(I + P_a D_a H_a)_{ki}|$ and $|H_{jk}|$ is contained in the determinants $|I + P_a D_a H_a|$ and $|H_a|$, it is possible to write $(K_a)_{ij}$ as the ratio of two polynomials, such that the denominator polynomial contains only the zeros of $|I + P_a D_a H_a| \, |H_a|$. If one then imposes the constraint that H contains no frequency-dependent elements, the system characteristic equation is given by

$$|I + P_a D_a H_a| = 0 \tag{19}$$

The nature of the roots of this equation is now investigated.

It is desirable at this point to make a distinction between the plant values used in the design of the system (the measured or supposed values) and the true physical values which are never known exactly. The supposed values are denoted with a prime (P_a', P_c') for this purpose. The difference between P' and P arises from errors (or limited precision) in measurement, nonlinearities, slight time variations in P, and other approximations known or unknown. For all practical purposes, P' is sufficiently close to P to be considered equivalent. However, in connection with stability considerations, a lack of awareness of the distinction between P' and P might cause the accidental cancellation of unstable-region poles and give a false stability picture. The prime notation is, therefore, employed to indicate plant values used for the design of the controller D. Similarly, the primes on K_a' and K_b' indicate that they represent the desired rather than the actual over-all system responses. One may write

$$D_a H_a = (P_a')^{-1} P_a' D_a H_a \tag{20}$$

or

$$(D_a H_a)_{kj} = \sum_{l=1}^{n} \frac{|P_{l'k}|}{|P_a'|} (P_a' D_a H_a)_{lj} \tag{21}$$

Since it is known from equation 3 that $P_a' D_a H_a$ is diagonal, then

$$(D_a H_a)_{kj} = \frac{|P_{jk'}|}{|P_a'|} N_{jj'} \tag{22}$$

where

$$N_{jj'} = (P_a' D_a H_a)_{jj} \tag{23}$$

$$(P_a D_a H_a)_{ij} = \sum_{k=1}^{n} P_{ik} (D_a H_a)_{kj} \tag{24}$$

$$(P_a D_a H_a)_{ij} = \sum_{k=1}^{n} P_{ik} \frac{|P_{jk'}|}{|P_a'|} N_{jj'} \tag{25}$$

$$(1 + P_a D_a H_a)_{ij} = \frac{\delta_{ij} |P_a'| + N_{jj'} \sum_{k=1}^{n} P_{ik} |P_{jk'}|}{|P_a'|} \tag{26}$$

Matrix $I + P_a D_a H_a$ is thus made up of two types of elements: type (a), equation 27, for all the elements along the principal diagonal; and type (b), equation 28, for all off-diagonal elements.

$$(1 + P_a D_a H_a)_{jj} = \frac{|P_a'| + N_{jj'} \sum_{k=1}^{n} P_{jk} |P_{jk'}|}{|P_a'|} \tag{27}$$

$$(1 + P_a D_a H_a)_{ij} = \frac{N_{jj'} \sum_{k=1}^{n} P_{ik} |P_{jk'}|}{|P_a'|}, \quad i \neq j \tag{28}$$

It is important to note

$$\lim_{P' \to P} \sum_{k=1}^{n} P_{jk} |P_{jk'}| = |P_a| \tag{29}$$

$$\lim_{P' \to P} \sum_{k=1}^{n} P_{ik} |P_{jk'}| = 0 \tag{30}$$

The determinant $|I + P_a D_a H_a|$ can now be written as

$$|I + P_a D_a H_a| = \prod_{j=1}^{n} \frac{|P_a'| + N_{jj'} \sum_{k=1}^{n} P_{jk} |P_{jk'}|}{|P_a'|} + W_0 \tag{31}$$

where W_0 represents the sum of all products of elements except the product of all the principal diagonal elements, which is shown separately.

Every term of W_0 contains at least two factors of type (b). If each of the terms making up W_0 is written as the ratio of two polynomials, it may be seen that their denominators are all identical, and equal to the denominator of the product term of equation 31 assuming that this product term is also written as the ratio of two polynomials. This follows from the definition of W_0 and the nature of the elements of $I + P_a D_a H_a$.

The numerators of the terms of W_0 are of a degree equal to or lower than the degree of the numerator of the product term of equation 31.

It follows from equation 30 and the foregoing discussion that W_0 has a negligible effect on the location of the zeros of $|I+P_aD_aH_a|$. It is, therefore, possible to write the characteristic equation, i.e., equation 19 in the following approximate form

$$\frac{|P_a'|+N_{jj}'\sum_{k=1}^{n}P_{jk}|P_{jk}'|}{|P_a'|}=0,\ j=1,2,\ldots,n \tag{32}$$

Applying equation 29 gives

$$\frac{|P_a'|+N_{jj}'|P_a''|}{|P_a'|}=0 \tag{33}$$

where

$$|P_a''|\approx|P_a'| \tag{34}$$

Now let

$$|P_a'|=\frac{A'}{B'}$$

and

$$|P_a''|=\frac{A''}{B''} \tag{35}$$

where the A's and B's are polynomials.

$$\frac{A'B''+A''B'N_{jj}'}{A'B''}=0 \tag{36}$$

It is assumed that N_{jj}' has no zeros in common with A' and no poles in common with B'. It is possible to factor equation 36 approximately into the following

$$A'=0 \tag{37}$$

$$B'=0 \tag{38}$$

$$1+(P_a'D_aH_a)_{jj}=0 \tag{39}$$

These three equations determine the normal modes of the system. For stability, none of them may have roots in an unstable region.

It follows from equations 37 and 38 that the elements of the plant matrix may have no poles and that the plant determinant $|P_a|$ may have no zeros which lie in an unstable region. The unstable region referred to is the right-half plane of the variable s for a continuous-data system, or the exterior of the unit circle of the z-plane for a sampled-data system. Poles or zeros on the boundary are not permissible.

Equation 39 can be recognized as the characteristic equation of the specified system. From equation 11 the specified over-all transfer matrix, which dictates the design of D_a, is given by

$$K_a'=(I+P_a'D_aH_a)^{-1}P_a'D_a \tag{40}$$

The corresponding characteristic equation is simply

$$(1+P_a'D_aH_a)_{jj}=0,\ j=1,2,\ldots,n \tag{41}$$

Equation 39 thus merely states that the over-all response which is specified must be stable if a stable system is to result.

The restriction on the location of the roots of equations 37 through 39 represents the necessary and sufficient condition for stability of a system in which $m=n$. The restriction on the poles applies to all elements of P_a' since a prohibited pole in any one of them will, of course, also show up as a pole of P_a'.

The only assumption made in equations 37 to 39 is that the values of the plant transfer functions used in the design of D_a are reasonably close to the actual values of the plant, i.e., $P'\approx P$.

Removal of Unstable-Region Poles and Zeros

In the foregoing it was assumed that none of the zeros of A' or poles of B' were common with zeros or poles of N_{jj}', which was assumed to be arbitrary. It is shown now that it is possible to achieve a stable system in spite of unstable-region zeros in $|P_a|$ or similar poles in P_a if such poles and zeros are common to N_{jj}'. This imposes a constraint on N_{jj}', and therefore also on the specified over-all response K'. Remembering the definition of W_0, it is possible to write

$$W_0=\frac{W_1}{|P_a'|^n} \tag{42}$$

Hence, from equation 31,

$$|I+P_aD_aH_a|=\frac{W_1+\pi_{j=1}^{n}\left(|P_a'|+N_{jj}'\sum_{k=1}^{n}P_{jk}|P_{jk}'|\right)}{|P_a'|^n} \tag{43}$$

Now consider $|P_a'|$ to have a zero $z=x$, and pole $z=y$ (the latter contributed by the element P_{rs}), both lying in an unstable region. Thus

$$|P_a|=\frac{z-x}{z-y}Y,\quad |P_a'|=\frac{z-x'}{z-y'}Y' \tag{44}$$

$$P_{rs}=\frac{U_{rs}}{z-y},\quad P_{rs}'=\frac{U_{rs}'}{z-y'} \tag{45}$$

$$|P_{jk}'|=\frac{Y_{jk}'}{z-y'}\text{ for }j\neq r\text{ and }k\neq s \tag{46}$$

$$|P_{jk}'|=|P_{jk}'|\text{ for }j=r\text{ or }k=s \tag{47}$$

Referring to equations 29 and 33 one can write now for equation 43

$$|I+P_aD_aH_a|=\frac{W_1+\pi_{j=1}^{n}(|P_a'|+N_{jj}'|P_a''|)}{|P_a|^n} \tag{48}$$

Substituting for $|P_a'|$ and $|P_a''|$ from equation 44 gives

$$|I+P_aD_aH_a|=\frac{W_1+\pi_{j=1}^{n}\left(\frac{z-x'}{z-y'}Y'+N_{jj}'\frac{z-x''}{z-y''}Y''\right)}{(Y')^n\frac{(z-x')^n}{(z-y')^n}} \tag{49}$$

It is seen now that $|I+P_aD_aH_a|$ has a zero of multiplicity n near $z=x$. If x is located in an unstable region, system instability can be avoided by letting the zero $z=x'$ be contained in N_{jj}' for all j. Let

$$N_{jj}'=(z-x)'N_{jj}'' \tag{50}$$

It is known from the definition of W_0 (and hence W_1) that W_1 consists of a sum of $n!-1$ products, each having n factors. These factors are either of type (a) or type (b) as given by equations 27 and 28, except that at least two factors in every product must be of type (b). This latter fact assures the vanishing of W_1 as $P'\to P$.

It follows now, since both $|P_a'|$ and N_{jj}' contain $z=x'$ as a zero, that W_1 contains $z=x'$ as a zero of multiplicity n. Let

$$W_1=(z-x')^nW_2 \tag{51}$$

Then equation 49 can be written as follows

$$|I+P_aD_aH_a|=\frac{W_2+\pi_{j=1}^{n}\left(\frac{Y'}{z-y'}+N_{jj}''\frac{z-x''}{z-y''}Y''\right)}{\frac{(Y')^n}{(z-y')^n}} \tag{52}$$

The unstable zero $z=x$ is no longer a root of this equation.

Use is now made of equations 45 through 47. Substitution of these equations into equation 52 gives

$$|I+P_aD_aH_a|=W_1+\left(\frac{Y'}{z-y'}+N_{rr}''\sum_{\substack{k=1\\k\neq s}}^{n}P_{rk}|P_{rk}'|+N_{rr}''\frac{U_{rs}}{z-y}|P_{rs}'|\right)\times\frac{}{\frac{(Y')^n}{(z-y')^n}}\times\pi_{\substack{j=1\\j\neq r}}^{n}\left(\frac{Y'}{z-y'}+N_{jj}''\sum_{\substack{k=1\\k\neq s}}^{n}P_{jk}\times\frac{|Y_{jk}|}{z-y'}+N_{jj}''P_{js}|P_{js}'|\right) \tag{53}$$

Both numerator and denominator are now multiplied by $(z-y)(z-y')^n$.

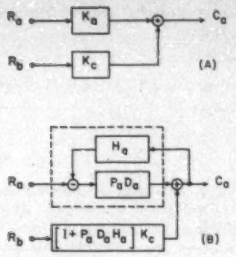


Fig. 3. Illustration of the effect of R_b on C_a

$$|I+P_aD_aH_a| = \frac{(z-y)(z-y')^nW_2}{(z-y)(Y')^n} + \left[(z-y)Y' + (z-y)(z-y')N_{rr}'' \sum_{\substack{k=1 \\ k\neq s}}^{n} \times \frac{P_{rk}|P_{rk}'| + (z-y')N_{rr}''U_{rs}|P_{rs}'|}{(z-y)(Y')^n} \right] \times \prod_{\substack{j=1 \\ j\neq r}}^{n} \left[Y' + N_{jj}'' \sum_{\substack{k=1 \\ k\neq j}}^{n} P_{jk}|Y_{jk}'| + (z-y')N_{jj}''P_{js}|P_{js}'| \right] \quad (54)$$

It is seen that the characteristic equation $|I+P_aD_aH_a|=0$ has a root in the neighborhood of $z=y$. This root can be removed by letting

$$N_{rr}'' = \frac{N_{rr}'''}{z-y'} \quad (55)$$

One obtains

$$|I+P_aD_aH_a| = \frac{(z-y)(z-y')^nW_2}{(z-y)(Y')^n} + \left[(z-y)Y' + (z-y)N_{rr}''' \sum_{\substack{k=1 \\ k\neq s}}^{n} \times \frac{P_{rk}|P_{rk}'| + N_{rr}'''U_{rs}|P_{rs}'|}{(z-y)(Y')^n} \right] \times \prod_{\substack{j=1 \\ j\neq r}}^{n} \left[Y' + N_{jj}'' \sum_{\substack{k=1 \\ k\neq j}}^{n} P_{jk}|Y_{jk}'| + N_{jj}''(z-y')P_{js}|P_{js}'| \right] \quad (56)$$

where W_2 has a first-order pole at $z=y$ and an nth-order pole at $z=y'$.

From equations 3, 11, 23, 50, and 55

$$N_{jj}' = \frac{z-x'}{z-y'} N_{jj}''' = (P_a'D_aH_a)_{jj} = \frac{(K_a'H_a)_{jj}}{1-(K_a'H_a)_{jj}} \quad (57)$$

One may thus conclude that unstable zeros in $|P_a'|$ may be removed by letting them be contained as zeros in $(K_a'H_a)_{jj}$ for $j=1, 2, \ldots, n$, and that unstable poles in an element P_{rs}' may be removed by letting them be contained as zeros in $1-(K_a'H_a)_{rr}$. This closely parallels the conclusions of Bergen and Ragazzini[2] and Bertram[3] for a 2-pole system.

The Case for $m \geq n$

The foregoing applied only to the case where $m=n$. It is shown now that a fairly simple extension is possible which covers the case where $m \geq n$. From equation 9,

$$C_a = K_aR_a + K_cR_b$$

Since C_a is to be a function only of R_a, K_c should be as small as possible. One may consider the output C_a as being made up of a "signal" K_aR_a and a "disturbance" K_cR_b. This is illustrated in Fig. 3(A). The system is shown in slightly different form in Fig. 3(B) to emphasize the closed-loop nature of K_a. It is seen that a large amplification in $P_aD_aH_a$ tends to minimize the effect of the disturbance. In a practical case the effect of R_b on C_a is small whether or not K_c is nearly zero, provided only that K_c is stable.

For the output C_a to be bounded for bounded inputs, both K_a and K_c must be stable. Since the conditions for the stability of K_a have already been determined, it is necessary to examine only K_c.

Solving equation 6 for D_c and substituting the result in equation 13 yields

$$K_c = (I+P_aD_aH_a)^{-1}[-P_a(P_a')^{-1}P_c'D_b + P_cD_b]H_b(I+D_bH_b)^{-1}H_b^{-1} \quad (58)$$

$$K_c = -(I+P_aD_aH_a)^{-1}P_a(P_a')^{-1}P_c'K_b + (I+P_aD_aH_a)^{-1}P_cK_b \quad (59)$$

The characteristic equation of K_c is given by the denominator of equation 59 when this equation is expressed as the ratio of polynomials, i.e., as the ratio of a matrix having only polynomial elements and a polynomial denominator. The roots of the characteristic equation are seen to be the following:

1. The zeros of $|I+P_aD_aH_a|$.
2. The zeros of $|P_a'|$.
3. The pole of P_c'.
4. The poles of P_c.
5. The poles of K_b.

In the foregoing, item 1 was examined already in connection with the stability of K_a. It was seen that none of the poles of P_a, none of the zeros of $|P_a|$, and none of the zeros of $1+(P_a'D_aH_a)_{jj}$ may lie in an unstable region.

The restriction implied by item 5 is of no consequence since K_b is entirely under the designer's control.

It is seen from equation 59 that unstable-region zeros of $|P_a'|$ need not cause instability since they may be cancelled by selecting a K_b' such that every element contains these zeros.

Poles of P_c are not known precisely and hence cannot be cancelled. A system having unstable-region poles in P_c, therefore, can not be stabilized. Physically, this follows from the fact that P_c is not enclosed within a feedback loop. A system designer should, of course, try to avoid this difficulty by arranging the control of P so that all elements having unstable-region poles are contained within P_a.

Illustrative Example

As an illustrative example, consider the following 3×2 plant transfer matrix:

$$P = \left\| \begin{array}{cc|c} \frac{2}{s+5} & \frac{2}{s+20} & \frac{1}{s+2} \\ \frac{1}{s-1} & \frac{1}{s+4} & \frac{2}{s+3} \end{array} \right\|$$

where the dotted line separates P_a and P_c.

The determinant $|P_a|$ is quickly found to be

$$|P_a| = \frac{20(s-4)}{(s-1)(s+4)(s+5)(s+20)}$$

It is seen that:

1. P_{21} has an unstable-region pole at $s=1$.
2. $|P_a|$ has an unstable-region zero at $s=4$, in addition to the unstable-region pole contributed by P_{21}.

According to the previous analysis, a stable system can only be obtained if all $(K_a'H_a)_{jj}$ contain $s-4$ as a factor, and $1-(K_a'H_a)_{22}$ contains $s-1$ as a factor. These two conditions are constraints imposed on the dynamic characteristics of the over-all response matrix K.

Consider the case for $j=2$. Let

$$(K_a'H_a)_{22} = (s-4)f_1(s), f_1(4) = \text{finite}$$

and

$$1-(K_a'H_a)_{22} = (s-1)f_2(s), f_2(1) = \text{finite}$$

Hence

$$1-(s-1)f_2(s) = (s-4)f_1(s)$$

where $f_1(s)$ and $f_2(s)$ are arbitrary functions of s meeting the constraint imposed by this relationship.

Let

$$f_2(s) = \frac{B}{s+a}$$

Then

$$(s-4)f_1(s)=1-\frac{(s-1)B}{s+a}$$

$$f_1(s)=\frac{(1-B)s+(a+B)}{(s-4)(s+a)}$$

Since $f_1(s)$ must be analytic at $s=4$, $s-4$ must be contained in the numerator in the foregoing; i.e., it must divide into the numerator without a remainder. Carrying out the long division and setting the remainder equal to zero gives:

$$a+4-3B=0$$

Hence

$$B=\frac{a+4}{3}$$

and

$$f_1(s)=-\frac{1+a}{3(s+a)}$$

$$(K_a'H_a)_{22}=-\frac{(1+a)(s-4)}{3(s+a)}$$

It is easy to verify that this expression for $(K_a'H_a)_{22}$ also meets the condition that $1-(K_a'H_a)_{22}$ have $s-1$ as a factor. The value of a in the foregoing is arbitrary and may be selected to meet performance requirements. (If a more complex form had been selected for $f_1(s)$, more arbitrary constants would be available.)

Conditions for Physical Realizability

For any stipulated system to be physically realizable, the normal response at any time can depend only on the past values of the input. In the case of sampled-data systems, this condition is met if the transfer function, when expanded as an infinite series in decreasing powers of z (where $z=e^{sT}$), has no term of higher power than z^0.

In the synthesis problem, plant P' is given, together with a stipulated over-all response K', and the transfer matrices D and H must be determined. One may assume that H will be designed solely on the basis of steady-state performance requirements and hence will involve only zero-memory elements. This eliminates any problem of physical realizability in connection with the matrix H.

The synthesis for the dynamic response of K is achieved through controller D. Solving equations 11 and 12 for D_a and D_b (and observing the rules with respect to the prime notation) gives

$$D_a=(P_a')^{-1}(I-K_a'H_a)^{-1}K_a' \tag{60}$$

$$D_b=(I-K_b'H_b)^{-1}K_b' \tag{61}$$

The matrix products $K_a'H_a$ and $K_b'H_b$ are both diagonal ($K_a'H_a$ only approximately). Hence $(I-K_a'H_a)^{-1}$ and $(I-K_b'H_b)^{-1}$ are also diagonal.

$$D_{ij}=\sum_{k=1}^{n}\frac{|P_{ki}'|}{|P_a'|}\frac{K_{kj}}{1-(K_a'H_a)_{kk}},\ i,j=1,2\ldots,n \tag{62}$$

$$D_{ij}=\frac{K_{ij}}{1-(K_b'H_b)_{ii}},\ i,j=n+1,n+2,\ldots,m \tag{63}$$

where

$$(K_a'H_a)_{kk}=\sum_{l=1}^{n}K_{kl}'H_{lk} \tag{64}$$

Equations 62 and 63 show that a physically realizable D_{ij} can always be achieved by selecting a K_{kj} which is of sufficiently low order in z. This imposes a slight restriction on the K elements, i.e., although the form of K is unrestricted, an unwanted delay may have to be added to achieve physical realizability. Thus given K_{kj}^0, it is always possible to achieve K_{kj}^1 where

$$K_{kj}^1=z^{-J}K_{kj}^0 \tag{65}$$

and s is some positive integer which depends on P, H, and K^0.

As an example, consider a plant having the following (sampled-data) transfer matrix:

$$P(z)=\left\|\begin{array}{c|c|c}\frac{5z^{-1}}{1-0.5z^{-1}} & \frac{3z^{-1}}{1-0.4z^{-1}} & \frac{2z^{-1}}{1-0.4z^{-1}}\\ \hline \frac{3z^{-1}}{1-0.4z^{-1}} & \frac{2z^{-1}}{1-0.5z^{-1}} & \frac{z^{-1}}{1-0.4z^{-1}}\end{array}\right\|$$

where the partitioning between P_a and P_b is indicated by the dotted line.

Calculating $|P_a|$ gives

$$|P_a|=\frac{z^{-2}(1+1.45z^{-1})(1-0.45z^{-1})}{(1-0.5z^{-1})^2(1-0.4z^{-1})^2}$$

Since $|P_a|$ is of order -2 in z, and all $|P_{ki}|$ are of order -1, the K_{kj} in equation 62 must be of order -1 or lower to satisfy the requirement for physical realizability. Since for practical considerations D_{ij} should be of order -1 (or lower), the K_{kj} are chosen to be of order -2 in z. Hence the K_{kj} must all have the following general form:

$$K_{kj}=z^{-2}K_{kj}^0$$

where K_{kj}^0 is a rational function in z^{-1} having neither a pole nor a zero at $z=0$.

Conclusions

The objective of the work presented here was to extend the synthesis procedure for multipole control systems described in a previous paper[4] so as to include consideration of some of the special problems of stability and physical realizability.

It is shown that a stable multipole control system can always be synthesized, subject to certain constraints imposed on the over-all response. If the plant has no elements with unstable-region poles and the plant determinant has no unstable-region zeros, the system may have any arbitrary response. However, if such poles or zeros are present, the over-all response becomes somewhat restricted. The restriction is generally not a serious one and the system designer retains wide latitude in his selection of a desired over-all response.

The synthesis method is also not seriously affected by the conditions of physical realizability which must be satisfied for the system controllers. With the insertion of a suitable delay, any arbitrary response characteristics can always be achieved.

Although the actual computational work required for the synthesis of multipole control systems is of necessity considerably more extensive than that required for a 2-pole system, it has been shown that the same techniques are applicable. It is thus possible to synthesize such systems in a straightforward manner without the necessity for trial-and-error studies with analog computers.

References

1. General Algebraic Method Applied to Control Analysis of Complex Engine Types, A. S. Boksenbom, R. Hood. *Technical Report 930*, National Advisory Committee for Aeronautics, Washington, D. C., Apr. 1949.

2. A Theory of Multidimensional Servo Systems, M. Golomb, E. Usdin. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 253, no. 1, Jan. 1952, pp. 29–57.

3. A Method for the Preliminary Synthesis of a Complex Multiloop Control System, J. Povejsil, A. M. Fuchs. *AIEE Transactions*, vol. 74, pt. II, July 1955, pp. 129–34.

4. A Synthesis Method for Multipole Control Systems, Herbert Freeman. *Ibid.*, vol. 76, pt. II, Mar. 1957, pp. 28–31.

5. The Analysis of Sampled-Data Systems, J. R. Ragazzini, L. A. Zadeh. *Ibid.*, vol. 71, pt. II, Nov. 1952, pp. 225–34.

6. Analysis and Synthesis of Sampled-Data Control Systems, Eliahu I. Jury. *Ibid.*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

7. The Theory of Pulse Monitored Servomechanisms and Their Use for Prediction, R. H. Barker. *Report No. 1046*, Ministry of Supply, Signals Research and Development Establishment, Christchurch, England, Nov. 1950.

8. Sampled-Data Processing Techniques for Feedback Control Systems, Arthur R. Bergen, John R. Ragazzini. *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 236–47.

9. Factors in the Design of Digital Controllers for Sampled-Data Feedback Systems, John E. Bertram. *Ibid.*, vol. 75, pt. II, July 1956, pp. 151–59.

10. The Synthesis of Multipole Control Systems, H. Freeman. *Doctoral Dissertation, Technical Report T-15/B*, Electronics Research Laboratory, Columbia University Engineering Center, New York, N. Y., Apr. 1956.

Electrical Control of Gas Turbines in Gas Transmission Service

R. A. YANNONE
ASSOCIATE MEMBER AIEE

A. B. NIEMOLLER
ASSOCIATE MEMBER AIEE

F. P. GOERTZEN
ASSOCIATE MEMBER AIEE

BY THE installation of new pipe lines and additional power the gas transmission industry continues to supply a greater percentage of the United States' fuel requirements. The newest type of plant for this service is the gas-turbine-driven direct-connected centrifugal compressor. This paper describes the electrical control for gas-turbine compressor stations on the Texas Eastern-Penn Jersey pipe line. Fig. 1 shows the Perulack station as built for one gas-turbine compressor unit.

Gas-Turbine Power Plant

The gas-turbine drive has proved itself in flexibility and adaptability to automatic control. Wider acceptance by the industry has led to its increase use as a prime mover for natural-gas compressor-drive service.

In May 1949, Westinghouse placed in service an 1,800-hp (horsepower) simple-open-cycle gas turbine, driving a direct-connected centrifugal compressor on a natural-gas line. This unit is in service today, having accumulated in excess of 43,000 hours of on-the-line operation.

Following the successful operating experience on this unit and the detailed studies of main natural-gas transmission-line requirements, larger units were produced. This paper discusses a unit rated 5,000 hp as used in gas transmission service.

Power Gas Cycle

The gas-turbine power plant is dual shaft and employs an open-type gas cycle with regenerator. The unit is rated 5,000 hp at 80 F (degrees Fahrenheit) and 1,000-foot elevation. The dual-shaft unit differs from the more common single-shaft unit in that the power turbine and its load are mechanically independent of the axial air compressor and its driving turbine. This provides increased flexibility over a wide speed range to handle line-flow variations. Fig. 2 shows a cut-away view of the gas-turbine power plant.

Axial Air Compressor

An 11-stage axial air compressor takes in atmospheric air and raises its pressure to approximately 4 atmospheres. The compressed air is passed through a diffuser where a portion of the kinetic energy of the air is recovered before passing on to the regenerator.

Regenerator

The regenerator or heat exchanger is used to recover heat from the power-turbine exhaust gas and to restore it to the compressed air entering the combustors. The addition of the regenerator results in an increase in thermal efficiency as it recovers a large percentage of the available exhaust heat. From the regenerator the air passes to the combustion chamber.

Combustion System

Combustion of the natural-gas fuel that supplies energy for the gas turbine takes place in the combustion chamber, the heart of the gas-turbine power plant. Scheduled fuel flow enters through fuel nozzles positioned at the entrance to each combustor, while compressed air enters the combustor casing from the regenerator. Within the chamber, six combustors or baskets of stainless steel are arranged radially between the compressor outlet and turbine inlet to contain the actual combustion process and to make possible the very high heat release.

Compressor and Power Turbine

Hot gases from the combustors expand through the 2-stage-compressor gas turbine which drives the axial air compressor. The gas leaves the compressor-driving turbine and further expands through a single-stage power gas turbine which drives the direct-connected centrifugal gas-pipe-line compressor. After leaving the power gas turbine, the exhaust gas passes through the regenerator and is then discharged to the atmosphere. This power gas cycle is diagrammed in Fig. 3. Fig. 4 is an installation view.

Auxiliaries

Starting Turbine

A 200-hp expansion gas turbine designed to operate from natural gas supplied from the pipe line is used to crank the axial compressor and turbine shaft. The expansion turbine is coupled to the auxiliary gear by a pneumatically operated clutch which is pressurized through a hollow shaft in this gear. The cranking power supplied by the expansion of gas accelerates the compressor and turbine shaft to firing speed and continues to assist until the combustion turbine unit is self-sustaining.

Turning Gear

The 2-hp a-c turning-gear motor is mounted on the auxiliary gear for rotation of the shaft after shutdown until the unit is cooled. The turning gear employs an overrunning centrifugal throw-out design which is automatically disengaged when the unit is started.

Lube-Oil Pumps

The main lubrication pump is a component of the turbine unit, driven from the auxiliary gear. Lubrication and control requirements for the gas turbine at idle speed and over are supplied by the main pump.

The 15-hp a-c motor-driven auxiliary lube pump is employed for starting the unit and has sufficient capacity for continuous operation in event of outage of the main pump.

The 3-hp d-c driven emergency pump operates from the station battery, in event of failure of the auxiliary pump or loss of a-c power during starting or shutdown. This pump can supply sufficient lubricating oil after the unit has shut down for adequate dissipation of heat conducted to the bearings during cool-down.

Lube-Oil Coolers

The lubricating-oil cooler fans are driven by 3-hp 2-speed a-c motors. Three lube-oil coolers are provided, each with capacity for removing 50% of the heat load.

Paper 57-991, recommended by the AIEE Petroleum Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Conference of the Petroleum Industry, Philadelphia, Pa., September 9–11, 1957. Manuscript submitted June 11, 1957; made available for printing August 16, 1957.

R. A. Yannone is with Westinghouse Electric Corporation, South Philadelphia, Pa., A. B. Niemoller is with Westinghouse Electric Corporation, Hillside, N. Y., and F. P. Goertzen is with Texas Eastern Transmission Corporation, Shreveport, La.

Fig. 1. Gas-compressor station at Perulack, Pa.

Ignition System

The ignition system employs spark plugs fired from an a-c transformer to ignite the air and gas mixture in two combustors. The remaining combustors are fired through the crossflame tubes which provide a cirumferential air connection between the six combustor baskets.

The spark plugs are mounted on retractable pistons which upon increased pressure in the combustor, due to self-sustaining combustion, operate to withdraw the plug from direct contact with the flame.

Controls

Turbine Control System

The control system is designed for automatically supervised operation over the entire speed-load range of the unit. In keeping with the trend toward automatic control systems in modern gas transmission stations, the unit employs an automatic sequence control capable of accelerating the turbine and gas compressor from rest to loading condition. The control possesses the capacity to check the sequence at determined key points. Should an event fail to occur in proper sequence, the control will act to shut down the unit. An automatic start is initiated by depressing the start button. No additional effort is required by the operator. The unit is automatically cranked for a prescribed purge period, fuel gas ignited, valves positioned and accelerated through the starting sequence to ready-to-load condition. The control requires 18 minutes to proceed through the entire sequence. Remote starting will be accomplished by the same control arrangement actuated by supervisory control.

The gas-turbine power plant control is basically an electrohydraulic system. Its purpose is to control and protect the gas turbine while maintaining the turbine variables of temperature and speed within the design limits. The automatic control has the capacity to do the following:

1. Start—the gas turbine and connected gas compressor are accelerated from rest to the ready-to-load condition.
2. Operate—maintain the gas turbine and connected gas compressor at the desired speed.
3. Stop—the unit will stop in automatic sequence, initiated by the operator or by any protective device.

Turbine Hydraulic Control

The governor system is designed for variable loads and speeds with changes in load accomplished by adjusting the setting of the variable-speed governors. To provide a speed signal, an impeller mounted on the rotating shaft develops an oil pressure that is proportional to the square of the turbine speed. Speed control of the gas turbine is accomplished in the same manner in which the conventional internal-combustion engine is controlled, namely, fuel is regulated by the speed governor. The quantity of fuel to the combustors to maintain speed and temperature between maximum and minimum limits is controlled by an integrated system composed of:

1. Axial-compressor speed governor—controls fuel flow during normal operation.
2. Acceleration limiter—limits the fuel flow during rapid speed changes.
3. Power-turbine speed limiter—limits the power turbine to rated speed.

The axial-compressor speed limiter and power-turbine speed limiter each receive a speed signal from their associated shaft-mounted impellers which in turn work on the forced-balanced governor. The output control pressure from the governor is utilized to position a displacement-type servomotor which, in turn, positions the throttle valve to control the fuel supply to the gas turbine. The control pressure output will be that required to balance the controller having the minimum set point. The governor is positioned to schedule fuel flow porportional to the control pressure, which is inversely proportional to the speed signal. As a result, the throttle closes as speed increases and opens as speed decreases. The speed setting of the governor can be changed by energizing a reversible motor through a switch mounted on the gas-turbine control panel.

Fig. 2. 5,000-hp gas-turbine power unit with cover removed

Turbine Electrical Control

The gas-turbine electrical control scheme consists of the necessary devices to sequence the start-up, operation, and shutdown of the units. Equipment employed in the control system is of failsafe design to the extent that failure of any control apparatus will result in shutdown.

The electrical control is designed to operate on d-c relaying to prevent shutdown of the turbine on momentary failure of alternating voltage. Power requirements for d-c relaying and control equipment are held to a minimum by simplified design. The d-c control provides maximum reliability.

The heart of the control circuitry is centered around the master control relay operated from the station battery. The relay is suitably interlocked to prevent turbine starting unless required conditions are normal, and serves as the center of the protective system.

The primary activators of the automatic control sequence are pressure switches which act to energize the clutch, ignition, and fuel-supply solenoids. Contacts of the pressure switches are also

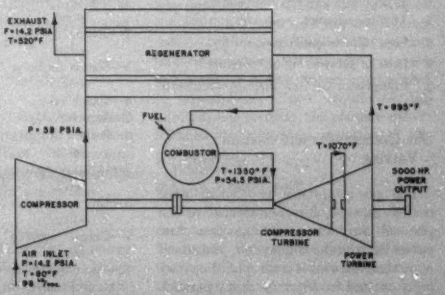


Fig. 3. Gas-turbine cycle diagram

Fig. 4 (left). 5,000-hp unit, gas-turbine end

Fig. 5 (right). 5,000-hp unit, gas - pipe - line compressor end

interlocked with the master relay which acts to shut down the unit in the event any operating condition is detected which could result in damage to the turbine.

Turbine Protective Devices

Protection against overspeed of the turbine shafts is provided by automatic-stop trips of the spring-loaded-weight type which are incorporated in the power-turbine and auxiliary gear shafts. At 110% speed, the centrifugal force on the plunger will overcome the spring load to permit the plunger to emerge from the shaft and strike the automatic-trip latch. Striking of the latch results in loss of overspeed pressure which acts to close the fuel isolation valve and shut down the unit.

The gas turbine is protected while starting or running by use of pressure and thermal switches which sense any abnormal condition and act to prevent a start or shut down the unit when such operation is detected. It is the function of the control system to supervise the operation of the protective devices which eliminate abnormal conditions that might result in equipment failure.

In summary, the control is designed to protect the turbine during operation against the following malfunctions:

1. Power-turbine overspeed.
2. Compressor-turbine overspeed.
3. Turbine exhaust overtemperature.
4. Loss of fuel-supply pressure.
5. Loss of lubricating-oil pressure.
6. Compressor-turbine inlet overtemperature.

Gas Compressor and Station Valving

The gas compressor is a single-stage centrifugal-type, direct-connected to the gas turbine. The total gas-line flow passes through the compressor, and where more than one unit is employed, they may be connected in series or series parallel. The installation shown in Fig. 5 is a typical centrifugal compressor with gas-turbine drive.

Gas-Compressor Auxiliaries

There are only three auxiliaries used with a centrifugal compressor: the bearing lubricating-oil pump, the lubricating-oil–to–air-heat exchanger fans, and the seal-oil pump. All other devices connected with the compressor are used for control, malfunction alarms, and protective devices. These devices provide protection against the following:

1. High bearing temperature.
2. Low bearing-oil pressure.
3. Low level in seal-oil reserve tank.
4. High discharge gas temperature.

Lubrication Oil

An a-c motor-driven lube-oil pump supplies oil to the bearings. Not only must proper oil pressure to the bearing be established, but the machine bearings should be at a minimum temperature of 68 F before the machine is rotated. To meet this requirement, thermostatically controlled heaters are immersed in the lube-oil reservoir to maintain a minimum oil temperature of 75 F.

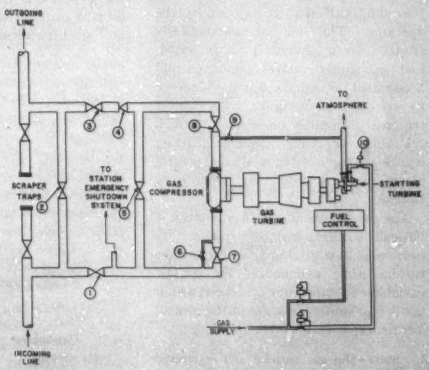
Seal Oil

A temperature-sensing device is located in the bearing-oil drain lines. When the bearing-oil drain temperature reaches 68 F and the proper oil pressure is being maintained in the supply header, a permissive circuit is completed which allows the next step in the sequence to occur, i.e., starting the seal-oil pump. This pump supplies the necessary high-pressure seal oil to provide an effective seal around the shaft between the high-pressure gas inside the case and atmosphere outside the case. An effective seal requires a differential pressure in the order of 2 to 25 pounds per square inch between the gas and the oil. This condition is maintained by a differential pressure regulator.

The seal-oil pump fills a high-pressure seal-oil reserve tank mounted on top of the compressor. The prime purpose of the reserve tank is to maintain seal-oil pressure during seal-oil pump failure or power outages. Gas pressure is taken

Fig. 6. Gas-piping diagram for single-unit station

1—Station suction valve
2—Station bypass valve
3—Station discharge valve
4—Check valve
5—Recycle valve
6—Pressurizing valve
7—Unit suction valve
8—Unit discharge valve
9—Case venting valve
10—Starting-turbine control valve



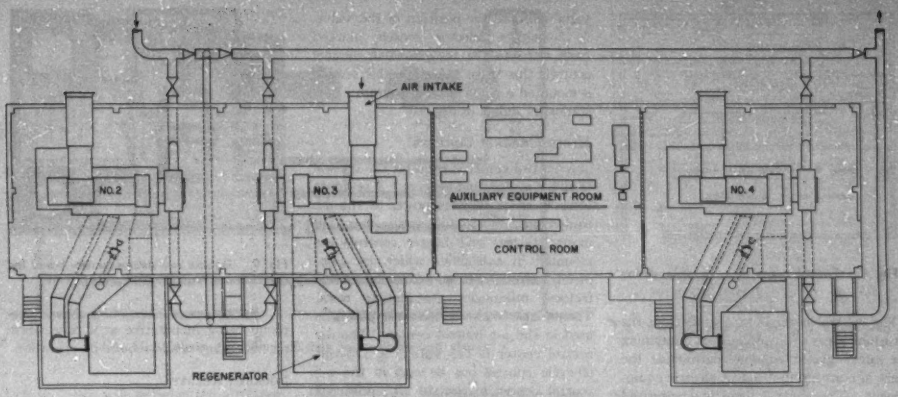


Fig. 7. Layout of a multiple-unit gas-turbine centrifugal compressor station

from the case to the top of seal-oil tank, this is called reference gas pressure. It can be seen that the differential seal-oil pressure, during seal-oil pump outages, is equal to the static head of the oil in the tank above the seal. The capacity of the seal-oil reserve tank is sufficient to allow operation of the unit for a minimum of 30 minutes. This provides ample time to shut down the unit without danger of damaging the seal element.

On the seal-oil reserve tank are mounted three level-sensitive control devices; one at the 100% full level, another at 85% full, and the third at the 50% full level. The level switches control the opening of the pressurizing valve and act to shut down the unit on low seal-oil level.

Gas Valves

To operate a gas transmission station, the following major gas valving per unit and station is generally supplied: station suction and discharge block valves, station discharge check valve, station bypass valve, station blowdown valve, recycle line and valve, and the scraper gate valves. Each gas compressor requires a unit suction valve, a discharge valve, and a bypass check valve. Fig. 6 shows typical piping for a gas-compressor unit.

Surge Detection

A centrifugal compressor is subject to surge at extremely reduced flow. It is desirable that these surges be quickly detected and corrective measures taken. When two or more compressors are operating and one unit is suddenly taken out of operation, the unit or units remaining on the line may have insufficient compression ratio to overcome the pressure differential existing across the station, and surge will result.

Surge can be detected by measuring the pressure drop across the station discharge check valve. A differential-pressure-sensing device, operating at full line pressure and capable of detecting a differential in terms of inches of water, is used to detect the surge condition.

Surge Control

When surging occurs, the gas-turbine speed is immediately reduced to minimum idle speed and the recycle-gas-valve motor is energized to open. The recycle valve then allows recirculation of gas around the station piping. This provides sufficient flow at the reduced speed to prevent surging.

When the pressure differential across the station has reached a reset value or equalized, the operator may accelerate the turbines to load the units. When any one of the turbine speed changers is energized to "raise," the recycle-valve motor is energized to close.

Station Emergency Shutdown

A station emergency shutdown system is used to protect the station in the event of an emergency condition. This system consists of two primary portions, the electrical section and the gas section.

Electrical Control

The electrical section derives its power from the 125-volt station battery. It is a closed-circuit system and therefore fail-safe. Once the circuit is de-energized, it cannot re-establish itself under any condition. The basic components are a solenoid-operated balanced-slide-block pilot valve to operate a pilot-gas-line trip valve, a pressure switch, several shut-down push-button stations, an auxiliary relay, a reset push button, and a system-reset indicating light.

Gas Control

The gas section consists of a power and pilot gas-supply tank pressurized from the main line. Power gas lines are required for operating the gas-powered hydraulically operated station suction, discharge, and blowdown valves. Pilot gas lines supply the manually operated pilot-gas-line trip valves, and the pilot-valve-operated trip valve.

Should an emergency condition arise, the pilot gas line is released to atmosphere by the operation of any push-button shut-down station or manual-trip-valve station. The release of pressure on the pilot gas line immediately initiates the closing of the station suction and discharge block valves and the opening of the station blowdown valve. All the gas contained in the station piping system is exhausted to atmosphere. Upon release of the pilot-gas-line pressure, the pressure switch initiates the shutdown sequence of the gas turbines.

Station Arrangement

Fig. 7 shows a possible arrangement of compressor units in a multiple station. To simplify the gas piping and ducting for the air intake and regenerator connections, the units are placed end to end.

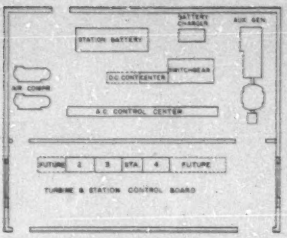


Fig. 8. Control room of a multiple station

The control room is separated from the turbine areas by firewalls, each containing a safety glazed window, permitting the use of conventional indoor electric equipment. The control room is pressurized with safe air from outside the building. Fig. 8 details the arrangement of the control and auxiliary equipment room.

Electrical Control

The design of the entire station has been engineered to provide a completely automatic system for start-up and shutdown sequences. Circuits are arranged to be failsafe. Protective devices are designed to indicate all troubles which might cause damage to the turbine or compressor equipment. These troubles are individually indicated on the annunciator. Whether the unit is shut down as a result of the trouble signal depends on the nature of the fault.

Gas-Turbine Control Board

Two sections of a dual switchboard contain the instrumentation, relays, and controls for each gas-turbine compressor unit. Each section is 28 inches wide by 90 inches high. The switchboard is 30 inches deep with hinged rear door on which the auxiliary control relays are mounted. The instruments, annunciator, and controls are mounted on the stationary front panels. The turbine control board is shown in Fig. 9.

Station Control Board

A single 40-inch-wide section of the dual switchboard contains instrumentation, relays, and control devices for station equipment. On this panel are indicated and recorded the station suction and discharge pressure, and the temperatures of incoming gas, discharge gas, and ambient air. A red and a green light in the symbol for each unit suction valve, unit discharge valve, and the station recycle valve indicate the position of the valve. A 2-position selector switch marked OPEN-AUTO in the recycle-valve symbol controls this valve. Also on this section is mounted a mimic piping diagram of the station as shown in Fig. 10.

Motor Control Centers

A control center is used for the control of all a-c motors and heaters. A separate control center supplies power to the d-c equipment. With this arrangement all motor starters and heater controls are mounted in assemblies which are completely wired to terminal blocks at the factory, minimizing installation work. Three-phase 440-volt 60-cycle power is used in the a-c control center. The d-c control center is 125 volts. A 115-volt 60-cycle control bus is used in the a-c control center to control all equipment included in the sequences.

Low-Voltage Switchgear

A small distribution switchboard controls both the normal 440-volt a-c public utility supply and an auxiliary generator. The 375-kva 3-phase auxiliary generator is driven by a reciprocating engine fueled from the line. On failure of the utility power, the auxiliary generator will start automatically and take over the entire a-c power load. The return to the public utility supply is effected by the operator.

Design Features

Operation Selector Switch

A 4-position selector switch marked OFF-PURGE-ACC-RUN is mounted on the turbine control board. For checking operation of the unit, this switch can be used to hold the starting sequence at the purging speed, in the accelerating condition blocking fuel and ignition, or in the shutdown condition. Normally this switch is left in the RUN position. In this position the starting sequence will proceed automatically to the ready-to-load condition when a start is initiated.

Maintenance Switches

Maintenance switches are provided for checking the operation of the equipment. These switches are 2- or 3-position depending on the particular application. A contact closed in the AUTO position of each of these switches is connected in series with the coils of the auxiliary master relays. This ensures that all maintenance switches are in the AUTO position during operation. The switches for the emergency lube-oil pumps are 2-position switches. The OFF position is omitted to eliminate the possibility of the pump being inoperative at a time when an emergency would require its operation.

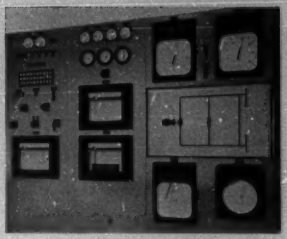


Fig. 9. Turbine and station control board for a single-unit station

Lockout Relays

A unique circuit, shown in Fig. 11, was developed to individually annunciate the troubles which operate the lockout relays to shut down the unit. Any trouble contact which closes energizes first the trip coil of the lockout relay. Contacts on the lockout relay then transfer the connection to the annunciator. Note that the only annunicator drop energized will be the one associated with the closed trouble contact.

Two trip lockout relays are used. Some of the trouble contacts indicate troubles which could be corrected by an operator. These trouble contacts are all connected to one of the lockout relays. The shutdown function of this relay is blocked when the starting selector switch is turned to the LOCAL position.

Timing Relays

Every effort was made in the development of the control scheme to keep the number of timing relays to a minimum. A total of three timing relays were used, each of which has two independently adjusted timing periods. These three timing relays accomplish six timing functions. A seventh timing function employs the opening time of the suction and discharge valves.

Speed Changer

A reversible d-c shunt motor is used to change the governor speed setting. During the starting sequence the operation of this motor is a part of the automatic sequence. At the completion of the starting sequence, the control of the speed-change motor is transferred to the operator.

The speed-changer motor is interlocked to block manual operation during the automatic sequences. Another inter-

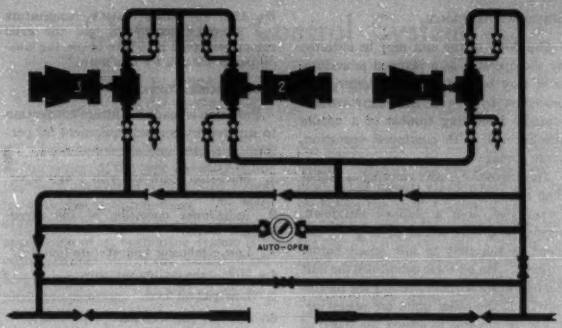


Fig. 10. Mimic gas-piping diagram for multiple-unit station

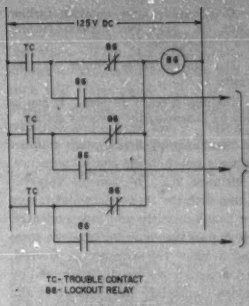


Fig. 11. Scheme of lockout relay and annunciator

lock prevents a decrease in the governor setting to a point below minimum idle speed.

During the shutdown sequence, the control of the speed-changer motor is transferred back to the automatic sequence, and the governor speed setting will return to ignition speed in anticipation of the next start.

During the automatic recycle sequence, the turbine must be returned immediately to the minimum idle speed. The station control solenoid valve accomplishes this function, overriding the speed-changer motor. The speed-changer motor automatically returns the governor setting to minimum idle speed in preparation to reload the unit at the completion of the recycle sequence.

Control Voltages

D-c control circuits from a battery supply have been universally accepted where maximum reliability is required. With the 125-volt battery already available, the choice of 125-volt direct current as the control medium was obvious.

Direct current is used for the operation of all control relays with the following exceptions: a-c timing relays with synchronous motors were used for all timing functions; contactors and relays within the a-c control center are operated from a 115-volt a-c control bus.

Instrumentation

Instrumentation is provided to give the operator a complete picture of the condition of each unit and of the station. In selection of the information to be displayed on the control board, special consideration was given to that information which would be helpful to the operator in anticipating and diagnosing troubles. Recorders were provided for those quantities essential to a history of operation.

Safety was a primary consideration in the transmission of data to the indicators and recorders. Pressure lines containing oil or gas are undesirable because a breakage would introduce either a combustible fluid or an explosive gas into the control room. Low-pressure air carried in instrument lines to pressure gauges on the control board is not hazardous and was used for a few of the indications. Electrical transmission of data is safe under all conditions and was used for major portions of the instruments.

Sequence Control

Purging

The minimum anticipated valve leakage will ensure a slight flow through the venting valve and will prevent contamination of the gas within the compressor. Purging of the gas compressor is included in the starting sequence by opening the pressurizing valve 60 seconds before the case venting valve is closed (steps 4 and 5, of the starting sequence).

The possibility of an explosive atmosphere in the gas-turbine components necessitates a complete purge during the starting cycle. This is the reason for the period of reduced flow to the cranking turbine after the first 20-second period of full flow to ensure breakaway and initial acceleration. The time required for the unit suction and unit discharge valves to attain the fully open position, steps 9(a) and (b) in the starting sequence, is used to time this purge period. When these two valves are fully open, a full flow of gas (step 11) is directed to the cranking turbine, and the turbine accelerates to the ignition speed.

Starting Sequence

The sequence of the operations in starting and placing a gas-turbine compressor unit in the ready-to-load condition is largely determined by the characteristics of the turbine, the compressor, the valves used in and the arrangement of the main piping. Operations which could damage equipment must be anticipated and measures taken to block their occurrence.

The sequence used at these stations is as follows:

1. The auxiliary master relays are energized and seal in.

2(a). Turbine auxiliary lube pump starts. This pump stops when sufficient pressure is developed by the main pump. (b). Compressor auxiliary lube pump starts. (c). Lube-oil cooler fans start.

3. Seal-oil pump starts.

4. Case pressurizing valve opens.

5. Case venting valve closes.

6. The master relay closes.

7. The cranking-turbine clutch valve is opened, admitting air to the clutch which engages.

8(a). The cranking-turbine shutoff solenoid valve is energized to allow a full flow of gas to the cranking turbine. (b). The unit discharge valve is energized to open. (c). The regenerator blowdown valves are reset.

9(a). The case pressurizing valve closes. (b). The unit suction valve is energized to open.

10. After 20 seconds of full flow of gas to the cranking turbine, the flow is reduced for purging. The purge period is approximately 5 minutes.

11. A full flow of gas is admitted to the cranking turbine.

12(a). The ignition circuit is energized. (b). Fuel is admitted to the combustors. (c). The speed-change motor is energized to increase the governor setting.

13(a). The ignition circuit is de-energized.

(b). Gas supply to the cranking turbine is shut off. (c). The cranking-turbine clutch is disengaged.

14. The speed-change motor is de-energized.

15. The unit is ready to load. It can be loaded by increasing the governor setting.

In the foregoing listing, each number represents a step in the control sequence. For instance, 2(a), (b), and (c) represent one step with the three operations taking place simultaneously and independently. Each step in the sequence must be completed to allow the sequence to proceed. The completion of a step is determined by a device completely independent of the equipment initiating the step.

The following is a brief explanation of the check which is made before the sequence can proceed. The numbering corresponds to the step numbers listed previously.

2. The starting sequence will be interrupted and the unit locked out unless lube pressure in both the turbine and compressor system has built up within 30 seconds.

3. The seal-oil pump will start only if the compressor auxiliary lube-pump output pressure is normal and the temperature of the oil is at least 68 F.

4. The pressurizing of the case will start only after a float switch on the seal-oil tank indicates 100% level.

7. Complete pressurizing of the case is required; it is indicated by a differential pressure switch.

8. Clutch operating pressure must be normal before this step can take place.

12. A pressure switch on the axial compressor output indicates ignition speed has been reached.

13. A pressure switch which receives oil pressure from an impeller on the axial compressor shaft signals maximum cranking speed.

14. A pressure switch which senses governor setting de-energizes the speed-changer motor.

Shutdown Sequence

Shutdown of the unit may be initiated by the operator as a planned procedure, or it may be initiated as an emergency procedure by one of the many trouble contacts indicating trouble of a nature which could, with continued operation, cause damage.

The tripout of a unit running at higher than idle speed will open the station recycle valve. For a planned shutdown, normal procedure would be to lower the speed of the unit to idle speed before tripping. Under these conditions the unit can be tripped without opening the recycle valve. If the other units are operating, it may be possible to adjust the speed of these units, keeping the station in operation while a unit is tripped out. Once tripped, the sequence is the same for all shutdowns and follows the following sequence:

1. The master and auxiliary master relays are de-energized.

2(a). The fuel valve is de-energized to close. (b). The unit suction valve is energized to close. (c). The speed-changer motor is energized to lower.

3. The unit discharge valve is energized to close.

4. The case venting valve is energized to open.

5. The seal-oil pump is de-energized.

6. The lube-oil pumps are de-energized.

Following are brief explanations of the checks made between points in the shutdown sequence:

3. The unit discharge valve starts to close after the suction valve has left its fully open position.

4. Both the discharge and suction valve must be fully closed.

5. Pressure in the compressor case must have reached zero.

6. After the exhaust stack temperature has decreased to 200 F and the axial compressor speed is 10% or below, the lube-oil pump will run for 5 minutes.

Supervisory Control

Provision is made for future connection to supervisory control equipment for performing control and alarm functions.

1. Start-stop operation with 3-way supervision.
2. Raise-lower operation of the speed-changer motor.
3. Low combustor temperature (outfire).
4. High exhaust temperature.
5. Low seal-oil level.
6. Prestart requirements (unit ready).
7. Lockout relay operation.
8. Annunciator operation.

For station functions:

1. Open-close of the recycle valve with 3-way supervision.
2. Closing of auxiliary generator circuit breaker.
3. Low battery voltage.

Provisions are made for continuous telemetering of the following conditions to the remote control point operator.

For each unit:

1. Speed of axial compressor turbine.
2. Speed of power turbine.

For station conditions:

1. Suction pressure.
2. Discharge pressure.

Conclusions

The gas-turbine-driven centrifugal gas-compressor station processes are readily adaptable to automatic sequence control and, where economics justify, to remote control.

Discussion

T. C. Heard (General Electric Company, Schenectady, N. Y.): The authors have compiled a clear word picture of the electrical control of this gas-turbine compressor station, a rather difficult thing to do without detailed diagrams. Due to the inherent automatic nature of gas turbines, the electrical and electrohydraulic control systems are of the utmost importance. The operational flexibility and reliability of a compressor station depends on these controls and basic turbine design.

The governing system of the gas turbine is the heart and brains of the compressor station. In addition to the speed-control governor mentioned in this paper, many gas turbines have additional governing to adjust automatically a nozzle between the axial-flow turbine and the load turbine, resulting in a wide range of speed and power at optimum efficiency. Also, control of the governor setting by pipe-line pressure signals is used for line pressure regulation.

For the electrically operated auxiliaries of most gas-pipe-line stations, 120/208- or 480-volt 3-phase power is used as indicated here. This station auxiliary power has sometimes been obtained by a generator, driven by (and mounted on) the gas turbine, but again with a stand-by power source.

Almost all gas-turbine control panels utilize d-c power for relay circuits as in this case. D-c control power has the inherent advantage of not being affected by the voltage fluctuations of a-c auxiliary power supply.

Provision for future supervisory control has been mentioned by the authors. The inherent automatic control of gas turbines allows very simple adaptation even when the original design did not consider this possibility. A few gas turbines are presently being operated by remote control. This is natural and relatively easy, and many expect that remote control of gas turbines will be effected often in the future.

Texas Eastern should be complimented on their important contributions in the centrifugal-compressor surge detection and control field.

Optimizing Control Systems

R. L. COSGRIFF
ASSOCIATE MEMBER AIEE

R. A. EMERLING
ASSOCIATE MEMBER AIEE

IT IS the purpose of an optimizing control system of the type discussed to adjust some variable of a controlled system for a maximum or minimum. The term optimizing is applied in that the usual application of an optimizing system is to adjust the operation of the controlled system for maximum efficiency, i.e. to optimize its operation.

Optimizing systems should not be confused with systems designed for optimum response such as systems with an optimum response for specific input signal, e.g., a step function or a periodic function. Optimizing systems, although satisfying the more general requirements of a servomechanism, differ materially from the more conventional-type systems. To illustrate this difference consider a conventional control system. Its desired output signal is a known function of the input signal. An optimizing system, however, does not have an input signal and the desired adjustment is known only after the controlled system is so adjusted that its output signal is a maximum or minimum. If the optimum adjustment changes with time it is necessary to continue adjusting the controlled system to ensure that the system is always correctly adjusted. An optimizing system is never exactly adjusted for optimum performance for any extended period, but a good control system will force the controlled system to be near the optimum adjustment at all times.

Several optimizing systems have been proposed. Many of these systems have been treated in some detail.[1] The optimizing system discussed in this paper was designed on the premise that control systems can be designed such that they analog the processes of the human operator in the manual control of the desired system.[2] A study was made to determine what decisions* are made by an operator in the manual adjustment of a system for the optimum condition, and these decisions were then analoged by use of logic or switching circuits.

Logic Circuits

Logic circuits[3] consist of three basic switching circuits, namely the AND, OR, NOT units. The input signals to and output signals from logic units are off-on in nature. Basically the AND unit has two or more input signals all of which must be ON before the output signal is ON. The OR unit has two or more input signals and the output will be ON if one or more of the input signals are ON. The NOT unit has one input signal and the output signal is OFF when the input is ON and vice versa. The representation for these basic logic units and one of many possible groups of electric logic circuits are shown in Fig. 1.

Design Requirements

To indicate the required logic circuits for an optimizing system observe the system configuration shown in Fig. 2. Here the motor position which adjusts the controlled system is represented by x and the resulting variable of the controlled system which is being adjusted for a maximum or minimum is represented by y. The logic circuits adjust the motor drive voltage e, such that the motor acceleration is always towards a position where y is a local maximum or a minimum. (In the following it is assumed that the goal of the optimizing system is to adjust y for a maximum.)

Three typical static characteristics of the controlled system are indicated in Fig. 3. Assume now that the initial conditions of x and y are given by point P_1. The initial value of $\dot{x}$ ($\dot{}$ is the differential operator d/dt) is taken as positive. Thus initially x will change such that y moves towards the maximum and $\dot{y}$ will also be positive. At P_2, the optimum position, $\dot{y}$ will change sign and the motor excitation e will also reverse its sign, reversing the motor acceleration; however, the inertia associated with the motor and load will carry x beyond the optimum position to the condition indicated by P_3. At P_3 the motor velocity will change from positive to negative and $\dot{y}$ should also change signs. Upon returning to state P_3 the motor excitation e will again change signs.

Consider the desired nature of the motor excitation. It will have two possible values. If e is ON, a positive voltage is applied to the motor terminals which in time will cause $\dot{x}>0$. If e is OFF, a negative voltage is applied to the motor terminals and in time $\dot{x}<0$. The signal e will be ON for the two conditions

$\dot{x}>0$ AND $\dot{y}>0$ (moving from P_1 to P_2, Fig. 3)

or

$\dot{x}<0$ AND $\dot{y}<0$ (moving from P_3 to P_1)

If the values of $\dot{y}$ and $\dot{x}$ are quantized such that the signal z_1 is ON when $\dot{y}>0$ and is OFF for $\dot{y}<0$, and z_2 is ON for $\dot{x}>0$ and OFF for $\dot{x}<0$, one can formulate the logic diagram shown in Fig. 4 for generating the signal e. There is one flaw in this logic. Consider now the condition at point P_1. Both $\dot{x}$ and $\dot{y}$ change signs at this point and as such both z_1 and z_2 change from OFF to the ON states. As a result, e should not change. In practice, however, two variables can not change instantaneously and as a result e will change to the OFF state at the time when z_1 or z_2 changes to the ON state, and since the motor velocity is nearly zero at this point the motor velocity will become negative. This process, when analyzed in detail and as has been shown in testing these control systems, causes the system to oscillate in the vicinity of P_1 rather than the desired position P_2. Likewise a similar problem arises at P_3. For this reason further logic units together with delay units must be added to the diagram to force e to remain OFF during the transition at P_3 and to remain ON during the transition at P_1. Circuits for performing these necessary steps are shown in Fig. 5 where the boxes labeled D are delay units (see also reference 4).

Analysis

The fundamental mode just described can be analyzed using the nonlinear-gain (describing-function) technique.[5] Essentially the problem is that of determining the magnitude of the steady-state oscillation of x about its optimum position, here assumed to be $x=0$ (see Fig. 3). It will be assumed that the input e to the motor is $\pm K$ and that the motor transfer function (see Fig. 2) is given by $G_m(p)$, i.e.,

$$G_m(p)e=x \qquad (1)$$

It is now assumed that x will be nearly sinusoidal and as a result it is only necessary to determine the fundamental amplitude and phase of the periodic component of e. If e changes sign at the optimum point, here assumed $x=0$, and there are

* These decisions are made during the learning process, but are not consciously made after the learning process.

Paper 57-778, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24-28, 1957. Manuscript submitted March 28, 1957; made available for printing October 10, 1957.

R. L. COSGRIFF is with Ohio State University, Columbus, Ohio, and R. A. EMERLING is with Systems Laboratory, Inc., Sherman Oaks, Calif.

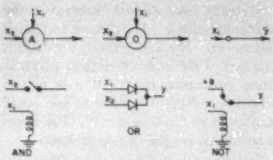


Fig. 1. Basic logic elements, their symbolic representation, and one set of electric circuits

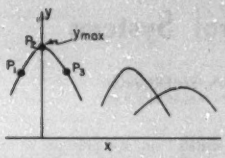


Fig. 3. Typical relationships between motor position x and output signal y of the controlled device

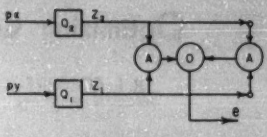


Fig. 4. Simplified logic network and quantizers Q_1 and Q_2

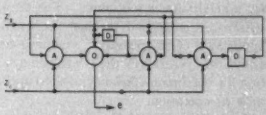


Fig. 5. Modification of logic network shown in Fig. 4 necessary to eliminate undesired modes of operation

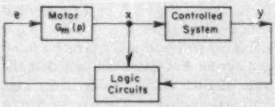


Fig. 2. Simplified block diagram of optimizing system

no delays in the controlled system or logic network, one can consider the controlled system and logic units as a feedback path from the motor shaft to the motor input. See Fig. 6 for nature of e, x, and y.

Let the fundamental component of x be $A\cos\omega t$ and the resulting fundamental component of e, namely e_1, becomes

$$e_1 = -\frac{4}{\pi} K \cos \omega t \tag{2}$$

Thus the gain of this feedback path can be represented by the ratio of the complex representation of e to the fundamental component of x, namely

$$\frac{E_1}{X} = G_1 = \frac{-\frac{4}{\pi}Ke^{j\omega t}}{Ae^{j\omega t}} = -\frac{4K}{\pi A} \tag{3}$$

where E_1 and X are the complex representations for the fundamental components of e and x. Note that $G_1X = E_1$ and $G_m(j\omega)E_1 = X$ has

$$G_m(j\omega)G_1X = X \tag{4}$$

as a representation for the system. Thus the amplitude and frequency of the oscillating state should be given by

$$G_m(j\omega)G_1 = 1$$

or

$$G_m(j\omega)^{-1} = G_1 = -\frac{4K}{\pi A} \tag{5}$$

The solution of equation 5 will give the approximate amplitude and frequency of the oscillation of x about its ideal value if such an oscillation exists. If $G(p)$ is of the form $G_0/(p^3\tau + p)$, no real solution for ω is possible in that $G(j\omega)$ is always complex while G_1 is real. Using more exact analytical techniques one finds that in this particular case x will converge in an oscillatory fashion toward $x = 0$ with an ever-increasing frequency and decreasing magnitude. Two assumptions have been made which are not met in practice, namely that $G_m(p)$ can be represented by a second-order differential equation and that there is no lag between the time $x = 0$ (py changes sign) and when e changes signs. If the switching circuits are made very fast, the motor can often be approximated by a transfer function of the form

$$G_m(p) = \frac{G_0}{p(m_0p^2 + m_1p + 1)} \tag{6}$$

In this case the two complex plots do cross and the values of A for G_1 and ω of $G_m(j\omega)$ at this intersection point indicate both the amplitude A and frequency ω of oscillation of x.

In fact one can demonstrate analytically that

$$\omega^2 = \frac{1}{m_2} \tag{7}$$

and that

$$A = \frac{4}{\pi} K \frac{G_0 m_2}{m_1} \tag{8}$$

It must be realized that the values of A and ω determined as indicated are not exact due to the approximations involved in assuming that x is sinusoidal.

Very often one cannot neglect the delays of the logic circuits. Likewise there is an effective delay which will be encountered due to the dead zone of the quantizers. If these delays are the same for x passing through zero in both directions, one can still determine an analytical expression for the amplitude and frequency of oscillation. Assume that in the vicinity of $x = 0$, y can be approximated by the curve

$$y = y_0 - \beta x^2 \tag{9}$$

where y_0 is the maximum value of y. Assuming that x is of the form $A\sin\omega t$, y becomes

$$y = y_0 - \beta\frac{A^2}{2}(1 - \cos 2\omega t) \tag{10}$$

and

$$\dot{p}y = -\beta A^2\omega \sin 2\omega t \tag{11}$$

Thus if the quantizer changes states at $py = -\Delta$ near point P_2, the delay time τ_q caused by the quantizer becomes

$$\tau_q = \frac{1}{2\omega}\sin^{-1}\frac{\Delta}{\omega\beta A^2} \tag{12}$$

By representing the delay in the logic circuits by τ_l, which is independent of A and ω, one finds that the fundamental component of e will be of the form

$$-\frac{4K}{\pi}\sin(\omega t - \omega\tau_l - \omega\tau_q) \tag{13}$$

As a result, the gain of the quantizer together with the switching circuits becomes

$$G_1 = \frac{-4K}{\pi A}\exp\left(-j\frac{1}{2}\sin^{-1}\frac{\Delta}{\omega\beta A^2}\right)\exp(-j\omega\tau_l) \tag{14}$$

Thus the solutions for A and ω can be determined from

$$G_m(j\omega)G_1 = 1$$

or

$$\frac{-4K}{\pi A}\exp\left(-j\frac{1}{2}\sin^{-1}\frac{\Delta}{\omega\beta A^2}\right) = G_m(j\omega)^{-1}\exp(j\omega\tau_l) \tag{15}$$

This last expression can be used to solve for both A and ω; however, it should be noted that the left side is a function of both ω and A. Because the variables ω and A are not separated, graphical solutions for A and ω become tedious. On the other hand, by letting $\omega A^2 = \kappa^2$, where κ is a new variable, one can separate the two variables κ and ω by expressing equation 15 in the following form

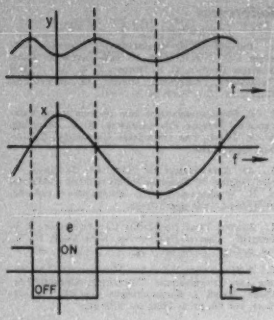


Fig. 6. Typical x, y, and e signals

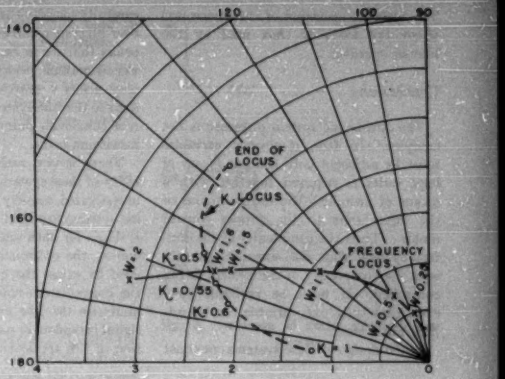


Fig. 7. Nyquist diagram used to determine amplitude and frequency of oscillation about the optimum condition

$$\frac{-4K}{\pi\kappa}\exp\left(-j\frac{1}{2}\sin^{-1}\frac{\Delta}{B\kappa^2}\right)=\frac{G_m(j\omega)^{-1}\exp(j\omega\tau_1)}{\omega^{1/2}} \quad (16)$$

Here the expression on the left side of the equality sign is now a function of κ alone and that on the right a function of ω. Such a change of variables makes possible solutions for κ and ω which are simply obtained by graphical methods as shown in the following example.

Consider a system characterized by $\beta=5$, $\Delta=1$, $K=1$, $G_m(i\omega)^{-1}=-\omega^2+j\omega$, and $\tau_1=0.1$. Plots of the right and left sides of equation 16 are shown in Fig. 7. Graphical solutions are then found to be $\kappa=0.535$ and $\omega=1.6$. By knowing κ and ω, the value of A can be determined, namely

$$A=0.535/\sqrt{\omega}=0.423$$

Performance Criterion

The degradation of an optimizing system's response from the ideal condition can generally be expressed as the average value of y_0-y. If the relationship between x and y is invariant and the deviation of y about y_{max} is small, one finds that the average value of y given by equation 10 is just

$$\bar{y}=y_0-\beta(A^2/2)$$

Thus the average value of y_0-y is given simply by

$$\beta(A^2/2)$$

This is normally referred to as hunting loss. The relative hunting loss of the system is defined by $(y_0-\bar{y})/y_0$ and is therefore given by

$$\beta(A^2/2y_0)$$

Multimode Operation

The system so far discussed can be considered the normal mode of operation of the optimizing system. However, by adding more logic circuits more modes of operation can be included to improve the over-all operation of the system. The first of these added modes can be considered a searching mode[4] which may be necessary if regions exist where y is nearly independent of x, namely dy/dx is sufficiently small that the quantizer Q_1 does not operate satisfactorily. Essentially this new mode of operation will cause x to traverse its full range, if necessary, in a cyclic fashion until a region where dy/dx has a reasonable value is reached. Once in such a region the system will be switched to the normal mode of operation.

If the optimum value of x changes very slowly in time compared with the time necessary for the system to seek out this optimum value, another mode of operation can be used. Essentially the logic circuits will detect the maximum value of x, namely x_M, and minimum value of x, namely x_m, when x is oscillating about is optimum value. Once x_m and x_M are determined the system is switched to this new mode and x is adjusted so that it will have a final value of

$$x=(x_m+x_M)/2$$

which in general will be very close to the optimum value of x. Once x has reached this value the optimizing system will be deactivated or switched to a second system which is to be adjusted. It is possible, for example, for such a system to tune several stages of a radio transmitter one at a time in cyclic fashion.

It is beyond the scope of this paper to consider the details of such multimodal operation.

Test Results

Two optimizing systems with a single mode of operation have been tested. The first[3] consisted of a logic network which was used to adjust a simulated motor and controlled system (the simulation was accomplished by use of an analog computer). Later a prototype system was constructed and used to adjust the position of a thermocouple above a heat source.[4] The former system operated in a very accurate fashion. In the later system the thermal eddies of air above the source caused the position of the highest temperature to change suddenly from one position to another. Likewise, there were large fluctuations in the temperature. The system would in general track the position with the highest temperature. Occasionally the changes of both temperature and position would cause the system to enter a region where the temperature was essentially uniform, and since the system was not able to detect any temperature changes with position, it would drift completely away from the desired region of operation. To determine how well the system was operating, the position of the thermocouple was adjusted manually using only the temperature indication as a guide. It was found that the operator had the same difficulty as the system in tracking the hottest position, and in general the machine appeared to be superior except that it did not have a searching mode of operation which the human operator used whenever he drifted away from the thermal source.

The speed of operation of the device was about the same as that used by the human operator.

Conclusions

The optimizing system discussed is not complex. In fact, if several variables must be adjusted only the one group of logic units is required together with a means of switching the output signal of the logic units to the proper control motor. Thus the complexity of the control system does not increase directly with the complexity of the system being controlled. It must be realized, however, that only one variable can be adjusted at a time.

In the design of these systems one must carefully consider the natural delays in the over-all system and expected nature of the peaks that are to be encountered. Likewise if multiple peaks occur one must employ either a searching mode or stored data so that the system will search out the true maximum point rather than a local maximum point.

There is one rather intriguing possibility of these systems which has not been investigated, namely, the use of optimizing systems to adjust conventional control systems for optimum response. For example, the optimizing system might be used to adjust the gain and damping of the conventional control system so as to minimize the rms error when the input signal is random in nature. An investigation of this application is now underway.

References

1. Principles of Optimalizing Control Systems and Application to the Internal Combustion Engine, C. S. Draper, Y. T. Li. *ASME Publication*, American Society of Mechanical Engineers, New York, N. Y., 1951.

2. A Servomechanism for Optimizing a Controlled System Performance, R. A. Emerling. *M. S. Thesis*, The Ohio State University, Columbus, Ohio, 1955.

3. An Application of Boolean Algebra to the Design of Electronic Switching Circuits, S. H. Washburn. *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 380–88.

4. Servos That Use Logic Can Optimize, R. L. Cosgriff. *Control Engineering*, New York, N. Y., vol. 2, Sept. 1955, pp. 133–35.

5. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, Ralph J. Kochenburger. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 270–84.

6. An Automatic Control System with Prevision for Scanning and Memory, N. H. Young. *Ibid.*, vol. 72, pt. I, Sept. 1953, pp. 392–95.

Compensating Saturation in Feedback Control Systems by Excess Error Storage

S. S. L. CHANG
MEMBER AIEE

R. W. ARCHBALD
ASSOCIATE MEMBER AIEE

BROADLY speaking, there are two types of saturation in a servo system: (1) sustained saturation, and (2) transient saturation. As an example of sustained saturation, the value of a ramp input may exceed the maximum speed of the servomotor when referred to identical basis. The only solution then is a trivial one, changing the servomotor. On the other hand, the transient saturation is by far more important and interesting technologically, and will be the subject of the present study.

A major cause of transient saturation is the use of lead networks, or lead-lag networks. While these networks are indispensable in improving speed of response as well as stability, they give rise to sharp peaks in the forcing function which are many times larger in magnitude compared to its normal value. As a result the controlled system is temporarily saturated. The system responds sluggishly and at times with large overshoots.

Fig. 1(A) illustrates the block diagram of a typical type 2 system with simple limiting. The transfer functions of the compensating network, G_c, and controlled, system, G_m, are given as

$$G_c = \frac{0.2(1+5.2s)}{1+0.2s} \tag{1}$$

$$G_m = \frac{2.5}{s^2} \tag{2}$$

The transient response of the system to a large step input is shown in Fig. 2(A). One notes that while the output function $f_1(t)$ from the compensating network is a very sharp peak exceeding 150 volts, its only effective portion is $f_2(t)$ which cuts off at 10 volts. As a result, the controlled system fails to reach the desired initial speed, and a large error tail is produced. To overcome this effect, an excess signal storage element is introduced into the system as shown in Fig. 1(B). The function of the storage element is to store any excess input signal above the saturation limit, and apply it to the controlled system immediately after the input signal falls below the saturation limit. Its input and output versus time characteristics are shown in Fig. 3. The restored signal area A_1 is slightly less than the clipped signal area A_2, by an amount depending on the signal magnitude as well as on the storage circuit design.

Fig. 2(B) illustrates the new system response subsequent to a step input of the same magnitude as that of Fig. 2(A). One notes that the effective forcing function $f_3(t)$ remains at a maximum value for sometime after $f_1(t)$ has reversed sign. The extended duration partially compensates for the lost magnitude, and consequently the system error is much reduced, as shown by the top oscillograph of Fig. 2(B).

An analog study of the afore-described simple method for compensating saturation was made, and it was found to be effective for a large variety of controlled systems. Typical among these are the systems mentioned and a highly oscillatory third-order system with the following transfer functions:

$$G_c = \frac{1.02[1+0.4s/\omega_1+(s/\omega_1)^2]}{1+0.615s/\omega_1+0.0756(s/\omega_1)^2} \tag{3}$$

$$G_m = \frac{1.273\omega_1}{s[1+0.4s/\omega_1+(s/\omega_1)^2]} \tag{4}$$

The system errors subsequent to step inputs of different magnitudes are illustrated in Figs. 4 and 5, respectively. In both figures, the oscillographs (A) to the left are for systems without storage elements, and the oscillographs (B) to the right are for systems with storage elements. The top graphs are for small inputs without saturation, and the middle and bottom graphs show increasing de-

Paper 58-92, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted October 7, 1957; made available for printing December 2, 1957.

S. S. L. Chang and R. W. Archbald are with New York University, New York, N. Y.

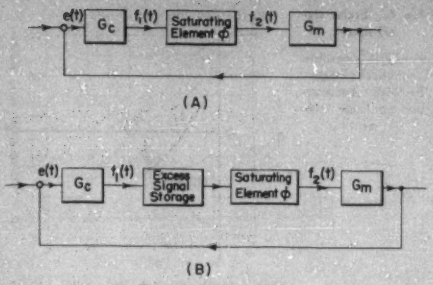


Fig. 1. Type 2 servo system

A—With simple limiting
B—With excess signal storage element

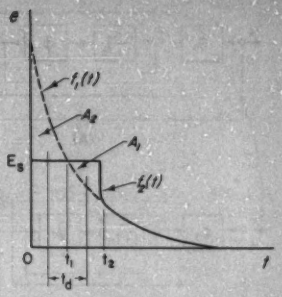


Fig. 3. Response of excess signal storage element

grees of saturation. The improvements in performance are substantial and without exception.

In general, it can be said that with the storage element, the range of acceptable system performance can be extended farther if the controlled system is less oscillatory. Fig. 5 illustrates the test result of a system with $\zeta=0.2$, and represents therefore a relatively poor example of compensation by this method.

In the following, two arrangements of excess signal storage making use mostly of existing components will be described. Theoretical grounds for the improvement, as well as criteria for large signal stability, will be discussed. It will become evident that this is a fairly general and effective method for compensating saturation at little added cost.

Excess Signal Storage Circuit

Figs. 6(A) and 6(B) represent two alternative arrangements for excess signal storage. In order to see its relation to the rest of the system the other system components are also shown, and the storage circuit is separated out by the broken lines. Fig. 6(A) illustrates the case when the output from the nonlinear element is conveniently accessible as a feedback signal. Fig. 6(B) illustrates the case when the output is not accessible or can be obtained only at appreciable cost. The nonlinear element Φ_1 is matched to the nonlinear element Φ, so that whenever Φ_1 saturates, Φ also saturates. With idealized saturation curves the two arrangements are identical in operating characteristics. Physically Φ_1 may represent the saturating limits of the amplifier, while the time constant T can be obtained by shunting a capacitor, C, across the input terminals of the amplifier K_c as shown in Fig. 7. While it is a normal practice to employ feedback amplifiers in servo systems, and the capacitor C bears a strong resemblance to the stabilizing capacitor, the foregoing arrangement is distinguished by the match-

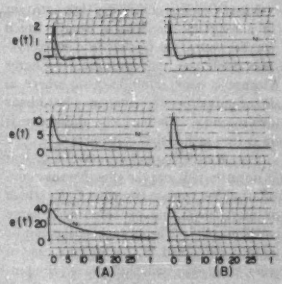


Fig. 4. Error of system described by equations 1 and 2

A—For increasing inputs
B—With excess signal storage element

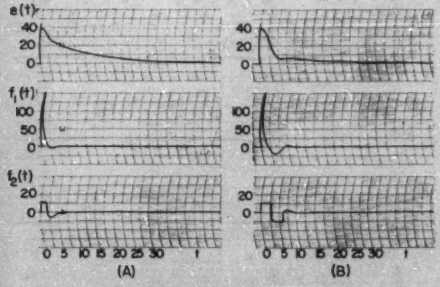


Fig. 2. Response of type 2 system

A—To a large step input
B—With excess signal storage element

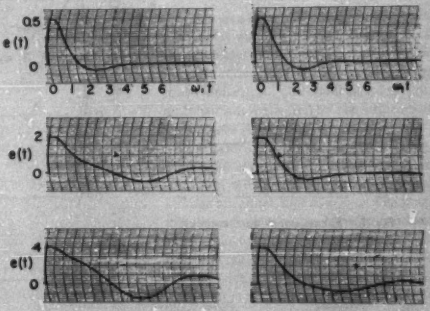


Fig. 5. Error of system described by equations 3 and 4

A—For increasing inputs
B—With excess signal storage element

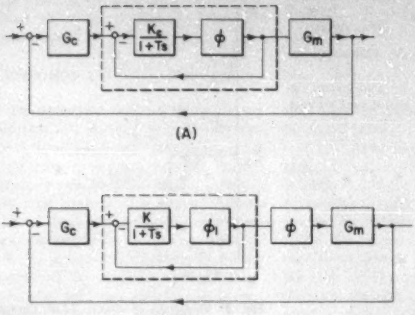


Fig. 6. Excess signal storage

A—With accessible nonlinear element
B—With unaccessible nonlinear element

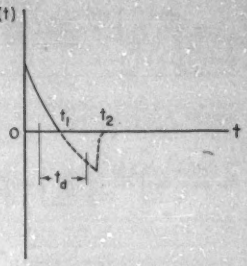


Fig. 9. Saturation error

ing of saturation limits and the large value of C. The stabilizing time constant of a feedback amplifier is usually much smaller compared to that of a servo system, while the time constant T for error storage is comparable to $1/\omega_0$, where ω_0 is the gain crossover frequency in radians per second. However, the difference is in design rather than hardware.

For relatively small inputs, the system is unsaturated, and the effective time constant of the storage circuit is T/K_C which is made negligible by selecting a large value for K_C. The storage action can be readily explained in the limiting case of very large gain such that the required input voltage e_i to the amplifier K_C is negligible, and the amplification μ is R_2/R_1. When the amplifier is operating in its linear range, $e_0 = -R_2/R_1 e_i$, and the currents i_1 and i_2 cancel out each other. When the input voltage e_i is large enough to saturate the amplifier, an excess current i_{ex} flows into C:

$$i_{ex} = \frac{e_i}{R_1} - \frac{E_s}{R_2} = \frac{1}{R_2}(\mu e_i - E_s) \qquad (5)$$

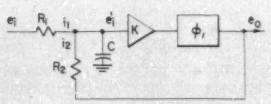


Fig. 7. Excess signal storage circuit

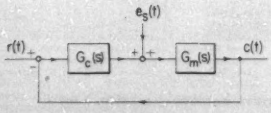


Fig. 8. Representation of systems shown in Fig. 1

The total charge accumulated is given by the following equation:

$$Q_c = \int_0^{t_1} i_{ex} dt = \frac{1}{R_2}\int_0^{t_1} (\mu e_i - E_s) dt \qquad (6)$$

Referring to Fig. 3, Q_c is simply A_2/R_2. After the input voltage drops below e_0/μ, the amplifier remains saturated for some time until the capacitor C is totally discharged. The discharge current is given by the following:

$$i_d = -(i_1 + i_2) = \frac{E_s}{R_2} - \frac{e_i}{R_1} = \frac{1}{R_1}(E_s - \mu e_i) \qquad (7)$$

The total discharge is

$$Q_d = \int_{t_2}^{t_3} i_d dt = \frac{1}{R_2}\int_{t_1}^{t_3} (E_s - \mu e_i) dt = \frac{A_1}{R_2} \qquad (8)$$

In case the time constant $R_1R_2C/(R_1+R_2)$ is much longer than t_3, Q_d would be approximately equal to Q_c. In general, Q_d is somewhat less than Q_c, and A_1 is less than A_2.

Analysis of System Response

It is customary to analyze a nonlinear system as a linear system superimposed with some extraneous inputs. The systems of Figs. 1(A) and 1(B) can be represented as the one shown in Fig. 8, where the saturation error $e_s(t)$ is

$$e_s(t) = f_2(t) - f_1(t) \qquad (9)$$

Corresponding to Fig. 3, the function $e_s(t)$ is a single pulse without the storage circuit and a dipolar pulse with the storage circuit as shown in Fig. 9. The solid

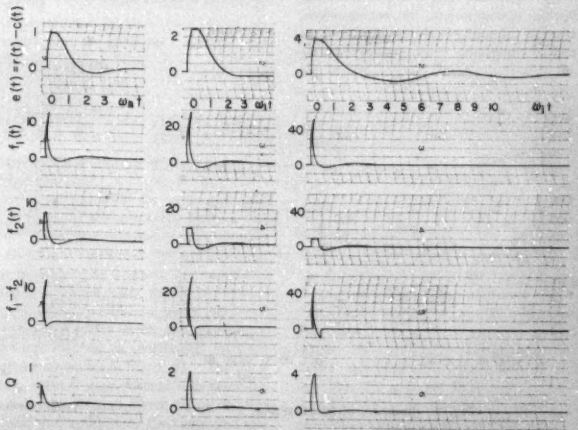


Fig. 10. Step input response of a system represented by Fig. 1(B) and equations 3 and 4

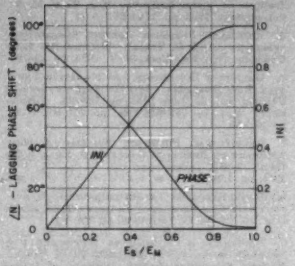


Fig. 11. Describing function of excess signal storage circuit

line represents $e_s(t)$ for both cases, while the broken line is a continuation of the solid line for the case with storage circuit only. As is well known, a single pulse has a uniformly distributed power spectral density up to a frequency of approximately $1/t_1$, while a dipolar pulse has a parabolic power spectral density up to a frequency of approximately $1/t_d$. For this reason the latter type of signal has relatively little power in the low-frequency region. Referring to Fig. 8, the Laplace transform of $c(t)$ can be written as

$$C(s) = \frac{G_c(s)\,G_m(s)}{1+G_c(s)\,G_m(s)}\,R(s) + \frac{G_m(s)}{1+G_c(s)\,G_m(s)}\,E_s(s)$$

$$= F(s)\,R(s) + \frac{F(s)}{G_c(s)}\,E_s(s) \qquad (10)$$

In equation 10, $F(s)$ is the closed-loop system function or control ratio, and $F(s)\,R(s)$ is the idealized linear response. With rare exceptions, deviation from the idealized linear response usually represents a change for the worse, and the term $F(s)\,E_s(s)/G_c(s)$ is to be made as small as possible. Normally $G_c(s)$ represents a lead network, resonant lead network, or a lead-lag network, and $1/G_c(s)$ has relatively high gain in the low-frequency range. As the storage circuit shifts the power spectral density of $e_s(t)$ into the higher frequency region, it is attenuated by $F(s)/G_c(s)$, which behaves like a low-pass filter.

From an alternative point of view, one may write:

$$\Delta C(s) = \frac{F(s)}{G(s)}\,A = \sum_i \frac{a_i}{s-\lambda_i} \qquad (11)$$

$$\Delta' C(s) = \frac{F(s)}{G(s)}\,A t_d s = \sum_i \frac{b_i}{s-\lambda_i} \qquad (12)$$

where λ_i's are roots of $F(s)/G(s)$, A is the clipped signal area, and t_d is the time delay in its restoration as shown in Fig. 3. $\Delta' C(s)$ and $\Delta C(s)$ represent deviations from linear response due to saturation with and without excess signal storage respectively. Equation 12 is derived on the assumptions that $A_1 = A_2 = A$ and $t_d << 1/\omega_c$. The Fourier coefficients b_i's are found to be much lower than a_i's in most cases.

Fig. 10 illustrates the step input response of a system represented by Fig. 1(B) and equations 3 and 4. The action of the excess signal storage circuit is increasingly apparent as the input amplitude is increased.

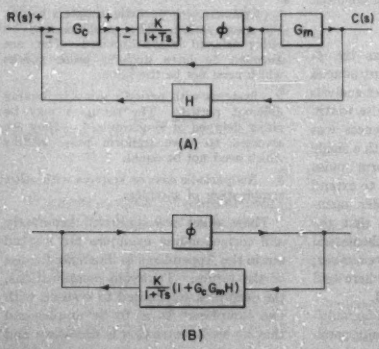


Fig. 12. System with excess signal storage element (A) and block diagram reduction of this system (B)

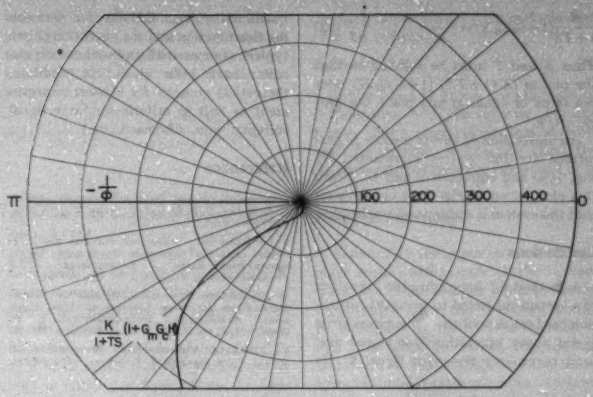


Fig. 13. Nyquist diagram of system represented by Fig. 1(B) and equations 3 and 4

Large Signal Stability

Large signal stability may be analyzed after obtaining the describing function of the nonlinear element or after using block diagram reductions which combine all the linear elements. The small signal system stability is not affected by the addition of the circuits shown in Figs. 6 and 7.

The describing function is given in Fig. 11 for an idealized circuit as shown in Fig. 7 in which the two areas of signal storage are equal, i.e., $A_1 = A_2$. The magnitude and phase of the transfer function are plotted against the ratio of the saturation voltage, E_s, to the maximum value of the sine wave input voltage, E_m. Large signals produce an added phase shift and decreased gain.

Nyquist diagram analysis of the system, Fig. 12(A), can be performed after using block diagram reductions to give Fig. 12(B). Instability is produced when the following obtains:

$$\frac{K}{1+TS}(1+G_cG_mH) = -\frac{1}{\Phi} \qquad (13)$$

Thus K and T must be selected so that the plot of $(K/1+TS)(1+G_cG_mH)$ does not cross or equal at any point the plot of $-1/\Phi$.

An illustration of this procedure is performed for the system represented by Fig. 1(B) and equations 3 and 4. From this plot shown in Fig. 13 it can be seen that this system is stable at all amplitudes.

Conclusions

A method for improving the operation of a system subjected to transient saturation has been described. The excess error signal above saturation was stored and used to prolong the error signal at its maximum value. Circuits for performing this function and the operation of two typical systems with and without the circuits, have been given. In addition, theoretical grounds for system improvement as well as criteria for large signal stability were discussed.

References

1. Principles of Servomechanisms (book), G. S. Brown, D. P. Campbell. John Wiley & Sons, Inc., New York, N. Y., 1948.

2. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, Ralph J. Kochenburger. *AIEE Transactions*, vol. 69, pt. 1, 1950, pp. 270–84.

3. Describing Function Method of Servomechanisms Analysis Applied to Most Commonly Encountered Nonlinearities, H. D. Graf. *Ibid.*, vol. 72, pt. II, Sept. 1953, pp. 243–48.

4. Feedback Theory—Some Properties of Signal Flow Graphs, S. J. Mason. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Sept. 1953, pp. 1144–56.

5. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955, pp. 88–100.

6. Harmonic Distortion and Negative Feedback, R. O. Rowlands. *Wireless Engineer*, London England, vol. 30, June 1953, pp. 133–35. Discussions: Sept., pp. 232–33; Oct., pp. 261–62; Nov., pp. 291–92.

7. Conditional Feedback Systems—A New Approach to Feedback Control, G. Lang, J. M. Ham. *AIEE Transactions*, vol. 74, pt. II, July 1955, pp. 152–61.

8. Block Diagram Transformations for Systems with One Nonlinear Element, T. M. Stout. *Ibid.*, vol. 75, pt. II, July 1956, pp. 130–41.

9. Nonlinear Compensating Networks for Feedback Systems, E. Mishkin, J. G. Truxal. *National Convention Record*, Institute of Radio Engineers, vol. 5, pt. 4, 1957, pp. 3–7.

10. Direct Synthesis Through Block Diagram Substitutions, O. J. M. Smith. *Ibid.*, pp. 8–16.

11. A Nonlinear Control System for Wide Range Input Signals, Julius Tou, Y. H. Ku. *Ibid.*, pp. 17–21.

Analysis of Multiple Sampler Systems With Finite Pulse Width, Open Loop

G. FARMANFARMA
STUDENT MEMBER AIEE

Synopsis: The Laplace transform method has been applied to the analysis of multiple-sampler systems. The method of analysis is based on a contour integral, which is derived from the Laplace transformation, and which mathematically describes the operation of a sampler with finite pulse width. This contour integral referred to as p-transformation in references 1 and 2 is employed for the exact analysis of linear multiple-sampler open-loop systems with finite pulse width. It is shown that by a combination of a few fundamental steps it is possible to treat any open-loop system. These steps have been outlined and simple examples given to demonstrate their application. The technique is general, no restriction being imposed on the number or the periodicity of samplers present. It may be applied to systems with delayed samplers, and periodic samplers with different periods and pulse widths, as well as to systems which include nonperiodic samplers both with respect to period and/or pulse width. In each case the solution is exact and in systems with periodic samplers it can always be expressed in a closed form. The limiting cases are also considered. As a consequence of these, a method is obtained for analysis of open-loop sample-data systems with zero pulse width which include two or more samplers.

Paper 57-1037, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted May 6, 1957; made available for printing September 30, 1957.

G. Farmanfarma is at the University of California, Berkeley, Calif.

The author wishes to express his gratitude to Dr. E. I. Jury for his help and encouragement during the preparation of this report. The research was supported by the Air Force Office of Scientific Research under Contract No. *AF 18(600)-1521*.

A METHOD referred to as the p-transform method was introduced in references 1 and 2 for the exact analysis of linear systems with finite pulse width. Its application in these references was limited to open-loop systems with a single periodic sampler having uniform pulse width. It is the purpose here to extend this analysis to multiple-sampler open-loop systems. It is assumed that the reader is familiar with the mathematical background given in these references; hence the details are omitted here and only a brief account is included.

A general study of the open-loop multiple-sampler systems leads to many possible cases. Among these are systems with delayed samplers, systems with periodic samplers which may or may not have the same periods and pulse widths, and systems in which both the pulse duration and the closing instants are nonperiodic. It has been found that all these cases may be analyzed by repeated application of a few fundamental steps. These steps have been outlined, and simple examples given to illustrate their applications.

Finite pulsed systems in general lead to more complicated equations than do the corresponding continuous systems. As a result the exact solution often seems relatively lengthy. However, tables are included which facilitate the relatively rapid solution of two sampler systems. These may be readily extended for use in more general cases.

The different cases arising have been divided into three categories:

1. Systems with periodic samplers having the same period. The samplers may be either delayed or synchronized. They are assumed to have uniform pulse widths which need not be the same.

2. Systems with periodic samplers having different periods. The samplers may be either delayed or synchronized. They are assumed to have uniform pulse widths which need not be equal.

3. Nonperiodic case or systems with other combination of samplers.

These cases are discussed separately and corresponding examples are worked out in the Appendixes to illustrate the use of the tables. To avoid complications, the examples are limited to systems with two samplers. It is to be emphasized that no such restriction is necessary and

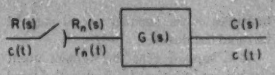


Fig. 1. Finite pulsed system with ideal sampler

that any systems may be treated, step by step, by the methods outlined.

The limiting cases are also investigated. When the pulse width $h \to 0$ and the sampler output is approximated by equivalent impulses, the corresponding sampled-data system results. As a consequence of this a method is obtained for analysis of open-loop sampled-data systems which have two or more samplers operating with different periods. The case where these samplers have a relative delay may also be treated.

Since the first publication of this material,[3] a further report[4] has been brought to the author's attention which analyzes sampled-data systems having periodic samplers with different periods. Such systems, in which the pulse width h is assumed to be zero, are discussed briefly in this paper. They are considered only as the limiting case of the more general finite pulsed systems whose analysis is the main object of this paper.

Mathematical Background

Fig. 1 shows an open-loop system with an ideal sampler. Assume that the sampler is arbitrarily closed at instants $T_0, T_1 \ldots T_n \ldots$, and opened at instants $(T_0+h_1) \ldots (T_n+h_{n+1}) \ldots$, etc. If $r(t)$ in Fig. 2(A) be the input to the sampler, the corresponding output is the time function $r_n(t)$ indicated in Fig. 2(B). Inspection shows that $r_n(t)$ may be written

$$r_n(t) = r(t)u_n(t) \tag{1}$$

where $u_n(t)$ is the function shown in Fig. 2(C). The Laplace transform of equation 2 may be evaluated as follows. Let

$$R_n(s) = \mathcal{L}[r_n(t)] \tag{2}$$

$$R(s) = \mathcal{L}[r(t)] \tag{3}$$

$$U_n(s) = \mathcal{L}[u_n(t)] = \sum_{n=0}^{\infty} \frac{e^{-T_n s}(1-e^{-h_{n+1}s})}{s} \tag{4}$$

When the complex convolution integral is applied,[1,2,5,6] $R_n(s)$ may be written

$$R_n(s) = \sum_{n=0}^{\infty} \frac{1}{2\pi j}\int_r R(v) \times \frac{e^{-T_n(s-v)}(1-e^{-h_{n+1}(s-v)})}{(s-v)}\,dv \tag{5}$$

where r is the path of integration enclosing all poles of $R(s)$ to its left. In the special case where the sampler is synchronized, i.e., $T_0=0$, and is periodic, the sampler output becomes[1,3]

$$R_p(s) = \frac{1}{2\pi j}\int_r R(v)\frac{(1-e^{-h(s-v)})}{(s-v)(1-e^{-T(s-v)})}\,dv \tag{6}$$

Equations 5, 6, and hence $R_n(s)$ and $R_p(s)$ may be evaluated by the residue method.[1,2,5,6] It is evident from the foregoing equations that the sampler may be regarded as an operator in the s-plane, transforming its input $R(s)$ according to equations 5 and 6 to give its output. For brevity, this will be referred to as p-transformation and denoted by the symbol P. It is to be noted that the p-transform is only a particular form of the Laplace transform.

The integrals of equations 5 and 6 may now be written in a more simplified form by the use of the symbol P. Thus, equation 5 becomes

$$R_n(s) = \sum_{n=0}^{\infty} \mathrm{P}_{T_n}^{T_n+h_{n+1}}[R(s)] \tag{7}$$

Similarly, equation 6 for the periodic case may be written

$$R_p(s) = \mathrm{P}_{h,T}[R(s)] \tag{8}$$

Tables I and II give a few examples of p-transforms and the inverse p-transforms. For a more comprehensive table and the method of obtaining them the reader is referred to either reference 1 or 2.

Delayed Functions

Essentially, the main interest at any stage is the relation between the Laplace transform of the sampler input to that of its output. When this is known, it may be multiplied by the transfer function of following block to obtain the Laplace transform of the output. In the linear systems under consideration, i.e., systems described by linear differential equations with constant coefficients, it is found that the primitive form of the input at any stage is limited to a function in the form of

$$C(s) = e^{-\sigma s}R(s) \tag{9}$$

where $\sigma \geq 0$ is a time delay measured in

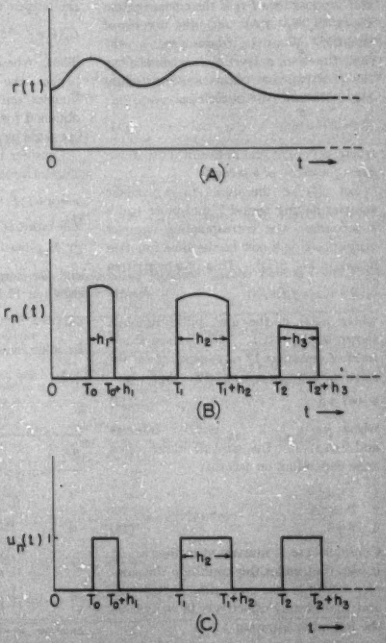


Fig. 2. Construction of sampler output through the multiplication of its input by pulse function $u_n(t)$

A—Input to sampler
B—Output of ideal sampler
C—Pulse function $u_n(t)$

Table I. Table of p-Transform Pairs

No.	C(s)	Cp(s) or Cn(s)
1	a_0	a_0
2	$\frac{a_0}{(s+\gamma)}$	$\left(\frac{1-e^{-h(s+\gamma)}}{1-e^{-T(s+\gamma)}}\right)\frac{a_0}{(s+\gamma)}$
3	$\frac{a_0 e^{\beta s}}{(s+\gamma)}$	$\left(\frac{1-e^{-h(s+\gamma)}}{1-e^{-T(s+\gamma)}}\right)\frac{a_0 e^{-\beta\gamma}}{(s+\gamma)}$
4	$\frac{a_1 s+a_0}{(s+\gamma)(s+\alpha)}$	$\frac{1}{(\alpha-\gamma)}\left(\frac{(a_0-a_1\gamma)(1-e^{-h(s+\gamma)})}{(s+\gamma)\ (1-e^{-T(s+\gamma)})}-\frac{(a_0-a_1\alpha)(1-e^{-h(s+\alpha)})}{(s+\alpha)\ (1-e^{-T(s+\alpha)})}\right)$
5	$\frac{e^{\beta s}(a_1 s+a_0)}{(s+\gamma)(s+\alpha)}$	$\frac{1}{(\alpha-\gamma)}\left(\frac{(a_0-a_1\gamma)e^{-\beta\gamma}(1-e^{-h(s+\gamma)})}{(s+\gamma)\ (1-e^{-T(s+\gamma)})}-\frac{(a_0-a_1)e^{-\beta\alpha}(1-e^{-h(s+\alpha)})}{(s+\alpha)\ (1-e^{-T(s+\alpha)})}\right)$
6	$P_T^{T+h}\left(\frac{a_0}{s+\gamma}\right)$	$(1-e^{-h(s+\gamma)})\frac{a_0 e^{-T(s+\gamma)}}{(s+\gamma)}$
7	$\sum_{n=0}^{\infty} P_{T_n}^{T_{n+1}+h_{n+1}}\left(\frac{a_0}{s+\gamma}\right)$	$\sum_{n=0}^{\infty} e^{-T_n(s+\gamma)}(1-e^{-h_{n+1}(s+\gamma)})\frac{a_0}{(s+\gamma)}$

seconds and $F(s)$ is a polynomial is s. All other forms of input may be obtained by a finite or infinite summation of this primitive form. Thus the first fundamental step is to obtain the sampler output assuming its input is the primitive function given by equation 9. With this known, the sampler output may then be obtained for any input by a summation process. Let

$$f(t)=\mathcal{L}^{-1}[F(s)] \tag{10}$$

and suppose that $f(t)$ is the time function shown in Fig. 3(A); $c(t)$, the inverse of equation 9 may be represented by the same function, delayed by σ seconds along the positive time axis as indicated in Fig. 3(B), and may be written

$$c(t)=d(-\sigma)f(t) \tag{11}$$

where $d(-\sigma)$ is the notation used to express the delay of σ seconds.

Let $c(t)$ be the input to a periodic sampler having period T and pulse width h seconds. The corresponding sampler output $c_p(t)$ will now be the time function shown in Fig. 3(D). This may be written

$$c_p(t)=d(-\sigma)f(t)u_p(t) \tag{12}$$

where $u_p(t)$ is the unit pulse function shown in Fig. 3(C). The Laplace transform of equation 12 now depends on the relation between σ, T, and h. Let

$$\sigma=nT+\beta \tag{13}$$

where $n=0,\ 1,\ 2\ \ldots$ positive integers and $0\leq\beta\leq T$. In general, three cases arise depending on whether

1. $h\leq\beta\leq T$
2. $0<\beta\leq h$
3. $\beta=0$ (14)

Consider case 1, illustrated in Fig. 3. It is seen that under this condition the sampler output is zero up to time $t=(n+1)T$, and that it may effectively be written as the following equation

$$c_p(t)=d(-\sigma)\{f(t)d[-(T-\beta)]u_p(t)\} \tag{15}$$

The Laplace transform of equation 15 is obtained by the complex convolution integral and is

$$C_p(s)=\frac{e^{-(n+1)Ts}}{2\pi j}\int_r F(v)e^{(T-\beta)v}\times\frac{(1-e^{-h(s-v)})}{(s-v)(1-e^{-T(s-v)})}\,dv \tag{16}$$

Equation 16 is similar in form to equation 6. Employing the same notations, the output may be written

$$C_p(s)=e^{-(n+1)Ts}P_{h,T}[F(s)e^{(T-\beta)s}] \tag{17}$$

Thus, when $h\leq\beta\leq T$, equation 17 expresses the Laplace transform of the sampler output, similar results may be obtained for the other two cases. These are given by relation 2 in Table III.

Consider the special cases where $\beta=T$. Then

$$\sigma=(n+1)T \tag{18}$$

The input is thus

$$C(s)=e^{-(n+1)Ts}F(s) \tag{19}$$

and the corresponding output given by equation 17 becomes

$$C_p(s)=e^{-(n+1)Ts}P_{h,T}[F(s)] \tag{20}$$

In other words, delays equal to integer multiples of the sampler period may be factored out directly.

Now, suppose that the sampler input $c(t)$ is a periodic time function with period T, equal to the sampler period. Then $C(s)$, the Laplace transform of the sampler input, is in general in the form of

$$C(s)=\frac{F(s)}{1-e^{-T(s+\alpha)}}=\sum_{n=0}^{\infty}e^{-nT(s+\alpha)}F(s) \tag{21A}$$

where all n's are positive integers. With a synchronized sampler assumed, and relation 20 utilized, the corresponding sampler output becomes

$$C_p(s)=\sum_{n=0}^{\infty}e^{-nT(s+\alpha)}\,P_{h,T}[F(s)]=\frac{1}{1-e^{-T(s+\alpha)}}\left[P_{h,T}[F(s)]\right] \tag{21B}$$

To generalize, this may be stated as follows; if the input to a synchronized, periodic sampler can be expressed in terms of delayed functions with delays equal to integer multiples of the sampler period T, so that the Laplace transform of the input is in the form

$$C(s)=\frac{F(s)}{(1-e^{-k_1T(s+\alpha_1)})(1-e^{-k_2T(s+\alpha_2)})\ldots(1-e^{-k_iT(s+\alpha_i)})} \tag{22A}$$

where all k_i's are integers and $F(s)$ is a polynomial in s. Then the corresponding sampler output may be written

$$C_p(s)=\frac{1}{(1-e^{-k_1T(s+\alpha_1)})(1-e^{-k_2T(s+\alpha_2)})\ldots(1-e^{-k_iT(s+\alpha_i)})}\,P_{h,T}[F(s)] \tag{22B}$$

Relations 21 and 22 are of utmost importance in the solution of multiple-sampler systems. Tables III and IV show additional input-output relations for delayed periodic inputs and delayed samplers.

Table II. Table of Inverse p-Transform Pairs

No.	Cp(s)	Conditions	Cp(t) $t=(n-1+m)T\quad t\geq 0$
1	$\frac{[1-ke^{-hs}]}{[1-e^{-T(s+\gamma)}]}\frac{a_0}{(s+\gamma)}$		$[1-kd(-h)]a_0 ke^{-\gamma t}$
2	$\frac{[1-ke^{-hs}]}{[1-e^{-T(s+q)}]}\frac{a_0}{(s+\gamma)}$	$q\neq\gamma$	$[1-kd(-h)]\left[\frac{a_0}{(e^{-T(q-\gamma)}-1)}(e^{\gamma T}e^{-T(nq+m\gamma)}-e^{-\gamma t})\right]$
3	$\frac{1-ke^{-hs}}{1-e^{-Ts}}\frac{a_1 s+a_0}{s(s+\gamma)}$		$[1-kd(-h)]\left[\frac{a_0 n}{\gamma}+\frac{(a_1\gamma-a_0)}{\gamma(e^{T\gamma}-1)}(e^{-T\gamma(m-1)}-e^{-\gamma t})\right]$
4	$\frac{1-ke^{-hs}}{1-e^{-T(s+\gamma)}}\frac{a_1 s+a_0}{s(s+\gamma)}$		$[1-kd(-h)]\left(\frac{a_0(1-e^{-n\gamma T})}{\gamma(1-e^{-\gamma T})}+\frac{(a_1\gamma-a_0)}{\gamma}ne^{-\gamma t}\right)$

Symbols: $d(-h)f(t)=u(t-h)f(t-h)$; $n=0,\ 1,\ 2\ldots+\infty$ integers; $0\leq m\leq 1$.

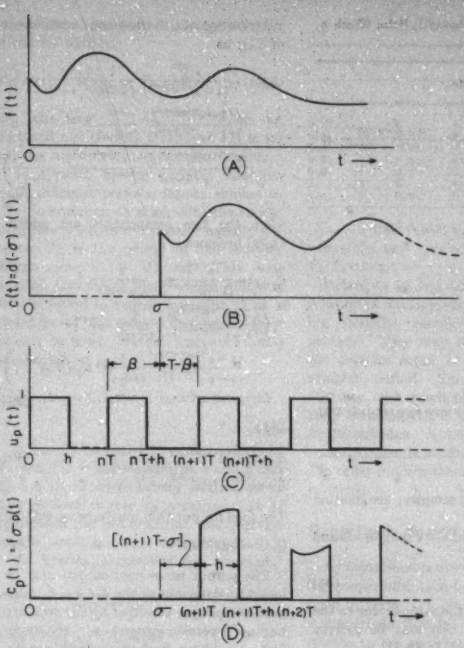


Fig. 3. Input-output relation for a delayed function

A—Arbitrary time function f(t)
B—f(t) delayed by σ seconds, considered as input to periodic sampler
C—Unit pulse function $u_p(t)$
D—Corresponding output of periodic sampler

Delayed Functions and Delayed Samplers

Consider the case where the sampler has a fixed operating delay of δ seconds with respect to the $t=0$ time axis, i.e., the first closure of the sampler occurs at $t=\delta$ seconds, after which it continues to operate normally. The effective result of such a fixed delay is to displace the unit pulse function $u_p(t)$ by δ seconds along the time axis.

Let the input to a periodic sampler be the primitive function

$$c(t)=d(t-\sigma)f(t) \tag{23}$$

Then, by constructions similar to Fig. 3, it can be shown that the corresponding sampler output is

$$c_p(t)=[d(t-\sigma)f(t)][d(t-\delta)u_p(t)] \tag{24}$$

assuming, in this particular case, that

$$\delta \geq \sigma \tag{25}$$

Then $c_p(t)$ may be written

$$c_p(t)=d(t-\sigma)(f(t)\{d[t-(\delta-\sigma)]u_p(t)\}) \tag{26}$$

The Laplace transform of equation 26 may be written

$$C_p(s)=\frac{e^{-\delta s}}{2\pi j}\int_r F(v)e^{(\delta-\sigma)v}\times \frac{(1-e^{-h(s-v)})}{(s-v)(1-e^{-T(s-v)})}\,dv \tag{27}$$

With the use of previous notations this becomes

$$C_p(s)=e^{-\delta s}\,P_{h,T}\,[F(s)e^{(\delta-\sigma)s}] \tag{28}$$

Equation 28 expresses the sampler output when $\delta \geq \sigma$. Other relations may be derived in a similar manner. Table IV shows the relation between input and output of a delayed, periodic sampler for various input functions.

When the sampler is not periodic, a step-by-step procedure must be used to obtain the output in the form of an infinite series. For example, Fig. 2 shows a function $c(t)$ being sampled by a non-periodic sampler with an operating delay of $\delta=T_0$ seconds. The sampler output in this case is expressed by equation 7. Similar relations may be obtained for other forms of input.

The foregoing describes the fundamental steps necessary for the solution of multiple-sampler systems. Repeated application of these steps may be used to solve any given open-loop system.

Periodic Samplers With the Same Periods

A discussion follows of systems having periodic samplers with the same period T. It is assumed that the samplers have uniform pulse widths h_1, h_2, etc., which need not be equal. Samplers with fixed delays are also considered.

A simple example can best illustrate the procedure. Fig. 4 shows a 2-sampler system. It is assumed that they have the same period T, but different pulse widths h_1 and h_2 respectively. In addition, the first sampler is assumed to have an operating delay of δ_1 seconds. Let

$$R(s)=\frac{1}{(s+\alpha)}$$

$$G_1(s)=\frac{1}{(s+\gamma)} \tag{29}$$

Here σ_1 is zero, and therefore $\delta_1 > \sigma_1$

Table III. Input-Output Relations for Synchronized Periodic Sampler Period T and Pulse Width h Seconds

No.	Input C(s)	Conditions	Output Cp(s)
1.	$F(s)$	$F(s)$ = a polynomial in s $n=0, 1, 2, \ldots$	$P_{h,T}[F(s)]=\frac{1}{2\pi j}\int_r F(v)\frac{(1-e^{-h(s-v)})}{(s-v)(1-e^{-T(s-v)})}\,dv$
2.	$e^{-(nT+\beta)s}F(s)$	$0\leq\beta\leq T$	
	(a)	$h\leq\beta\leq T$	$e^{-(n+1)Ts}\,P_{h,T}\,[F(s)e^{(T-\beta)s}]$
	(b)	$0<\beta\leq h$	$[e^{-(nT+\beta)s}\,P_{h^1-\beta}\,[F(s)]+e^{-(n+1)Ts}\,P_{h,T}\,[F(s)e^{(T-\beta)s}]]$
	(c)	$\beta=0$ $n=0, 1, 2, \ldots$	$e^{-nTs}\,P_{h,T}\,[F(s)]$
3.	$\frac{e^{-(nT+\beta)s}}{[1-e^{-kT(s+\alpha)}]}F(s)$	$k=0, 1, 2, \ldots$ $0\leq\beta\leq T$	
	(a)	$h\leq\beta\leq T$	$\frac{e^{-(n+1)Ts}}{(1-e^{-kT(s+\alpha)})}\,P_{h,T}\,[F(s)e^{(T-\beta)s}]$
	(b)	$0<\beta\leq h$	$\frac{1}{(1-e^{-kT(s+\alpha)})}[e^{-(nT+\beta)s}\,P_{h^1-\beta}[F(s)]+e^{-(n+1)Ts}\,P_{h,T}\,[F(s)e^{(T-\beta)s}]]$
	(c)	$\beta=0$	$\frac{e^{-nTs}}{(1-e^{-kT(s+\alpha)})}\,P_{h,T}\,[F(s)]$

Table IV. Input-Output Relations for Delayed Periodic Sampler with Period T, Pulse Width h, and Delay of δ Seconds

No.	Input C(s)	Conditions	Output Cp(s)
1	$F(s)$	$F(s)$ = a polynomial in s	$e^{-\delta s}\ \mathrm{P}_{h,T}\ [F(s)e^{\beta s}] = \frac{e^{-\beta s}}{2\pi j}\int_{\gamma} F(y)e^{\beta y}\,\frac{[1-e^{-h(\alpha-\gamma)}]}{(s-y)[1-e^{-T(\alpha-\gamma)}]}\,dy$	
2	$e^{-\sigma s}F(s)$	$\sigma \geq \delta$	
		$n = 0, 1, 2, \ldots$	
	Let $(\sigma-\delta) = nT+\beta \ldots$	$0 \leq \beta \leq T$	
	(a)	$h \leq \beta \leq T$	$e^{-(\delta+(n+1)T)s}\ \mathrm{P}_{h,T}\ [F(s)e^{(T-\beta)s}]$
	(b)	$0 < \beta \leq h$	$[e^{-\sigma s}\ \mathrm{P}_h{}^{-\beta}[F(s)]+e^{-(\delta+(n+1)T)s}\ \mathrm{P}_{h,T}\ [F(s)e^{(T-\beta)s}]]$
	(c)	$\beta = 0$	$e^{-\delta s}\ \mathrm{P}_{h,T}\ [F(s)]$
3	$e^{-\sigma s}F(s)$	$\sigma \leq \delta$	$e^{-\delta s}\ \mathrm{P}_{h,T}\ [F(s)e^{(\delta-\sigma)s}]$
4	$\frac{e^{-\sigma s}}{(1-e^{-kT(s+\alpha)})}\,F(s)\ldots$	$\sigma \geq \delta$	
		$k = 0, 1, 2, \ldots$	
		$n = 0, 1, 2, \ldots$	
	Let $(\sigma-\delta) = nT+\beta \ldots$	$0 \leq \beta \leq T$	
	(a)	$h \leq \beta \leq T$	$\frac{e^{-(\delta+(n+1)T)s}}{(1-e^{-kT(s+\alpha)})}\ \mathrm{P}_{h,T}\ [F(s)e^{(T-\beta)s}]$
	(b)	$0 < \beta \leq h$	$\frac{1}{(1-e^{-kT(s+\alpha)})}\ [e^{-\sigma s}\ \mathrm{P}_h{}^{-\beta}\ [F(s)]+ e^{-(\delta+(n+1)T)s}\ \mathrm{P}_{h,T}\ [F(s)e^{(T-\beta)s}]]$
	(c)	$\beta = 0$	$\frac{e^{-\sigma s}}{[1-e^{-kT(s+\alpha)}]}\ \mathrm{P}_{h,T}[F(s)]$

Thus the output of the first sampler is given by equation 28 as

$$R_{p_1}(s) = e^{-h_1 s}\ \mathrm{P}_{h_1,T}\left(\frac{e^{\delta_1 s}}{(s+\alpha)}\right) \tag{30}$$

$R_{p_1}(s)$ may now be obtained from relation 3 of Table I and is

$$R_{p_1}(s) = e^{-\delta_1 s}\,\frac{e^{-\delta_1\alpha}(1-e^{-h_1(s+\alpha)})}{(s+\alpha)(1-e^{-T(s+\alpha)})} \tag{31}$$

Therefore $C(s)$ becomes

$$C(s) = R_{p_1}(s)G_1(s) = e^{-\delta_1 s}\times \frac{e^{-\delta_1\alpha}(1-e^{-h_1(s+\alpha)})}{(1-e^{-T(s+\alpha)})}\,\frac{1}{(s+\alpha)(s+\gamma)} \tag{32}$$

which may be written

$$C(s) = \left(\frac{e^{-\delta_1(s+\alpha)}}{1-e^{-T(s+\alpha)}}\,\frac{1}{(s+\alpha)(s+\gamma)} - \frac{e^{-(h_1+\delta_1)(s+\alpha)}}{1-e^{-T(s+\alpha)}}\,\frac{1}{(s+\alpha)(s+\gamma)}\right) \tag{33}$$

Note that $C(s)$ is composed of two periodic functions, both with period T but one having a delay δ_1 and the other a delay of $(h_1+\delta_1)$ seconds.

$C(s)$ is the input to the second sampler which is assumed to have no operating delay. The output $C_{p_2}(s)$ of this sampler now depends on the relation between the delays, δ_1 $h_1+\delta_1$, and the pulse width h_2, and period T of the sampler; see relation 3 of Table III.

Now, suppose that $\delta_1 < T$ (i.e., $n_1 = 0$ and $\beta_1 = \delta_1$)

and $\delta_1 > h_2$ (34)

Then the output of the sampler due to the first component of $C(s)$ may be written from relation 3(A) of Table III as

$$C_{p_{21}}(s) = \frac{e^{-\delta_1\alpha}e^{-T\delta}}{(1-e^{-T(s+\alpha)})}\times \mathrm{P}_{h_2,T}\left(\frac{e^{(T-\delta_1)s}}{(s+\alpha)(s+\gamma)}\right) \tag{35}$$

$C_{p_{21}}(s)$ can now be obtained from relation 5 of Table I, and is

$$C_{p_{21}}(s) = \frac{e^{-T\delta}}{[1-e^{-T(s+\alpha)}]}\,\frac{1}{(\gamma-\alpha)}\times \left(\frac{e^{-T\alpha}(1-e^{-h_2(s+\alpha)})}{[1-e^{-T(s+\alpha)}]}\,\frac{1}{(s+\alpha)} - \frac{e^{-T\gamma}e^{-\delta_1(\alpha-\gamma)}(1-e^{-h_2(s+\gamma)})}{[1-e^{-T(s+\gamma)}](s+\gamma)}\right) \tag{36}$$

Also assume that

$$h_1+\delta_1 = T+\beta_2, \text{ i.e., } n_2 = 1,\ \beta_2 = (h_1+\delta_1-T) \tag{37}$$

and that $\beta_2 > h_2$. Then, relations 3(A) of Table III together with relation 5 of Table I may again be used to express the contribution due to the second component of $C(s)$ as

$$C_{p_{22}}(s) = \frac{e^{-2T\delta}}{(1-e^{-T(s+\alpha)})}\,\frac{1}{(\gamma-\alpha)}\times \left(\frac{(1-e^{-h_2(s+\alpha)})}{(1-e^{-T(s+\alpha)})}\,\frac{1}{(s+\alpha)} - \frac{e^{-(h_1+\delta_1)(\alpha-\gamma)}(1-e^{-h_2(s+\gamma)})}{(1-e^{-T(s+\gamma)})}\,\frac{1}{(s+\gamma)}\right) \tag{38}$$

when the two components are added, $C_{p_2}(s)$ becomes

$$C_{p_2}(s) = \frac{e^{-T\delta}}{[1-e^{-T(s+\alpha)}]}\,\frac{1}{(\gamma-\alpha)}\times \left(\frac{(1-e^{-h_2(s+\alpha)})}{(1-e^{-T(s+\alpha)})}\,\frac{1}{(s+\alpha)}\,(e^{-T\alpha}+e^{-T\delta}) - \frac{(1-e^{-h_2(s+\gamma)})}{(1-e^{-T(s+\gamma)})}\,\frac{1}{(s+\gamma)}\,(K_1+K_2e^{-T\delta})\right) \tag{39}$$

where

$$K_1 = e^{-T\gamma}e^{-\delta_1(\alpha-\gamma)} \tag{40}$$

$$K_2 = e^{-(h_2+\delta_1)(\alpha-\gamma)}$$

The final output $Q(s)$ may now be written

$$Q(s) = C_{p_2}(s)G_2(s) \tag{41}$$

The output time function $q(t)$ may be found by determining the inverse p-transform of equation 41, either by the ordinary Laplace inversion integral or by the modified z-transform method outlined in references 1 and 2. Note that if a third sampler were present $Q_p(s)$ could be found by the use of relation 22 and Tables III and IV.

Periodic Samplers With Different Periods

Systems with periodic samplers having different periods are investigated in this section. It is assumed that the samplers have uniform pulse widths, which need not be equal. The effect of samplers with operating delays may also be considered.

It has been shown that if $C(s)$, the input to a synchronized periodic sampler with period T_1, is of the form

$$C(s) = \frac{1}{1-e^{-kT_1(s+\alpha)}}\,F(s) \tag{42}$$

where k is an integer, and $F(s)$ a poly-

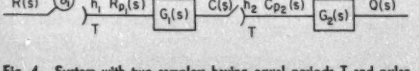


Fig. 4. System with two samplers having equal periods T and pulse widths h_1, h_2, with first sampler delayed by δ_1 seconds

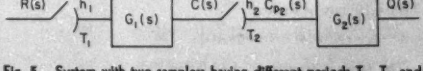


Fig. 5. System with two samplers having different periods T_1, T_2, and pulse widths h_1, h_2

nomial in s, the output of the sampler could be written

$$C_p(s) = \frac{1}{1-e^{-kT_1(s+\alpha)}} \, P_{k,T_1}[F(s)] \qquad (43)$$

i.e., the term $1/(1-e^{-kT_1(s+\alpha)})$ can be factored out directly whenever kT_1 is an integer multiple of the sampler period.

In systems having periodic samplers with different periods, inputs similar to $C(s)$ in equation 42 arise, with the exception that T_1 does not usually equally T_2, where T_2 is the period of the sampler whose output is desired. This can greatly complicate the problem unless a way is found to express equation 42 as a function of T_2 only. The latter may readily be done, once two integers b and q are found, so that

$$\frac{T_1}{T_2} = \frac{b}{q} \quad (b, q = 1, 2, 3, \ldots \text{ positive integers}) \qquad (44)$$

It should be noted that, for any given T_1 and T_2, a pair of integers b and q can always be found to satisfy relation 44 to any desired degree of accuracy. It is preferable, but not essential, that q be the lowest possible integer denominator.

To illustrate, consider the function

$$C(s) = \frac{1}{1-e^{-T_1(s+\alpha)}} \, F(s) \qquad (45)$$

which may be written as

$$C(s) = F(s) \sum_{i=0,1,2}^{\infty} e^{-iT_1(s+\alpha)} \qquad (46)$$

If equation 44 is substituted in equation 46, $C(s)$ may be written

$$C(s) = F(s) \sum_{i=0}^{\infty} e^{-i\frac{b}{q}T_2(s+\alpha)} \qquad (47)$$

Equation 47 may be expanded, factored, and written

$$C(s) = F(s)\left[1+(1+e^{-bT_2(s+\alpha)}+ e^{-2bT_2(s+\alpha)}+\ldots)\sum_{i=1}^{q} e^{-i\frac{b}{q}T_2(s+\alpha)}\right] \qquad (48)$$

which reduces to

$$C(s) = \frac{F(s)}{1-e^{-T_1(s+\alpha)}} = \frac{F(s)\left(1+\sum_{i=1}^{q-1} e^{-i\frac{b}{q}T_2(s+\alpha)}\right)}{(1-e^{-bT_2(s+\alpha)})} \qquad (49)$$

Similarly, it may be shown that

$$C(s) = \frac{F(s)}{(1-e^{-T_1(s+\alpha)})^r} = \frac{F(s)\left(1+\sum_{i=1}^{q-1} e^{-i\frac{b}{q}T_2(s+\alpha)}\right)^r}{(1-e^{-bT_2(s+\alpha)})^r} \qquad (50)$$

where b, q, and r are positive integers and $T_1/T_2 = b/q$.

Relations 49 and 50 are important in solution of multiple-sampler systems having periodic samplers with different periods. They may be used to express the sampler input in terms of T_2, the sampler period. Tables III and IV together with relations 21 and 22 may then be employed in finding the output of the sampler. A numerical example of such a system is worked out in Appendix I in order to show the application of these relations.

Nonperiodic Systems

As mentioned previously, systems having nonperiodic samplers must be treated term by term. In general, the output of a nonperiodic sampler is in the form of an infinite series, as is seen from Fig. 2 and equation 7. This complicates the treatment of systems with two or more nonperiodic samplers, for then the output is expressed in terms of two or more infinite series. In mixed systems containing both periodic and nonperiodic samplers, Tables III and IV can still be usefully employed. An illustrative example of such a system is given in Appendix II.

Limiting Cases in Multiple-Sampler Systems

Consider an open-loop system with a single periodic sampler. Let $R(s)$ be the input to this sampler. Then $C_p(s)$, the final output, may in general be written

$$C_p(s) = R_p(s)G(s) \qquad (51)$$

where

$$R_p(s) = P_{h,T}[R(s)] \qquad (52)$$

It is interesting to determine the limits of $C_p(s)$ as h is allowed to approach certain limiting values. There are four possible limiting cases arising in such a system. All of these are discussed in references 1 and 2. However, the two cases of interest here are:

1. The limit as h approaches the sampler period T.

Table V. Sample Calculation for Evaluating Equation 71

			a	b		f	g	k	c(t)
i	n	mT_1	$(1-m)T_1$	$e^{-i(1-m)kT_1}$	e^{-i}	$a-b$	$\frac{a-b}{1.72}$	$n-f$	$d(-0.6)^g$	$(g-k)$
0	1	0	1	2.72	1	1.72	1	0	0	0
0.2	1	0.2	0.8	2.22	0.82	1.40	0.82	0.18	0	0.18
0.4	1	0.4	0.6	1.82	0.67	1.15	0.67	0.33	0	0.33
0.6	1	0.6	0.4	1.50	0.55	0.95	0.55	0.45	0	0.45
0.8	1	0.8	0.2	1.22	0.45	0.77	0.45	0.55	0.18	0.37
1	2	0	1	2.72	0.37	2.35	1.37	0.63	0.33	0.30
1.2	2	0.2	0.8	2.22	0.30	1.92	1.12	0.88	0.45	0.43

2. The limit as h becomes so small that the unit pulse function, Fig. 3(C), may be replaced by a unit impulse function which has infinite amplitude, zero pulse width, and unit area.

References 1 and 2 show that in the first case the limit of $C_p(s)$ is

$$\lim_{h \to T} C_p(s) = C(s) = R(s)G(s) \qquad (53)$$

which corresponds to the Laplace transform of the continuous output, a result to be expected from physical considerations. It was also demonstrated that in the second case

$$\lim_{\substack{h \to 0 \\ k \to \infty \\ hk \to 1}} R_p(s) = R(s)|_{z=e^{sT}} \qquad (54)$$

where k is the amplitude of $u_p(t)$, and $R(z)$ the z-transform of $R(s)$. Consequently the limit of the output becomes

$$\lim_{\substack{h \to 0 \\ k \to \infty \\ hk \to 1}} C_p(s) = R(z)G(s) \qquad (55)$$

which is also a known result used in sampled-data systems, where the pulse width is considered to be extremely small.

In multiple-sampler systems these limiting cases are of course still valid. However, in such systems it is not always possible to derive the limiting conditions directly from the p-transform of the final output. For example, referring to Fig. 5, it cannot always be asserted that

$$\lim_{\substack{h_1 \to T_1 \\ h_2 \to T_2}} C_{p_2}(s)G_2(s) = R(s)G_1(s)G_2(s) \qquad (56)$$

The reason for this is better appreciated when it is realized that the nature of $C_{p_2}(s)$ depends on whether $h_1 \geq h_2$ or $h_1 \leq h_2$. Thus, with $C_{p_2}(s)$ once determined, on the assumption that $h_1 \geq h_2$, under no circumstances can h_1 be allowed to become less than h_2. However, it is possible that the initial values for a certain system are such that $h_1 > h_2$ but the limiting conditions demand that $h_1 < h_2$. Under these conditions difficulties arise which do not allow the derivation of the limiting condition

directly from the p-transform of the final output. In view of this it is generally advisable and more convenient to derive the limiting conditions directly from the system after inserting the desired limits in the corresponding samplers. To illustrate this procedure, the following limiting conditions will be derived from the system shown in Fig. 5.

1. The limit as $h_1 \to T_1$ and $h_2 \to T_2$.
2. The limit as $h_1 \to 0$ and $h_2 \to T_2$.
3. The limit as $h_1 \to 0$ and $h_2 \to 0$.

In cases 2 and 3 it is understood that $u_p(t)$ then approaches the unit impulse function.

CASE 1

When $h_1 \to T_1$ and $h_2 \to T_2$, the samplers may be replaced by a line representing direct through transmission. It is readily seen from Fig. 5 that under these conditions

$$Q(s) = R(s)G_1(s)G_2(s) \tag{57}$$

which is the Laplace transform of the continuous system.

CASE 2

Under the conditions of case 2 the output becomes

$$Q(s) = R_{T_1}(z)G_1(s)G_2(s) \tag{58}$$

where $R_{T_1}(z)$ is the z-transform of $R(s)$ with $T = T_1$. The inverse time function $q(t)$ for such a system may be determined in closed form, and as a continuous function of time by the use of the inverse p-transform tables[1,2] or the modified z-transform[7] tables.

CASE 3

Referring to Fig. 5, let

$$R(s) = \frac{1}{s} \tag{59}$$

Under the assumed conditions the output of the first stage $C(s)$ may be written

$$C(s) = z_{T_1}[R(s)]G_1(s) = \frac{1}{1-e^{-T_1 s}}G_1(s) \tag{60}$$

and the final output becomes

$$Q(s) = z_{T_2}[C(s)]G_2(s) \tag{61}$$

To find $Q(s)$ it is therefore necessary to obtain the z-transform of equation 60 with respect to a period T_2 different from T_1. Relation 49 may be used to achieve this in the manner: Let the ratio of T_1 to T_2 be

$$\frac{T_1}{T_2} = \frac{b}{q} \tag{62}$$

where b and q are positive integers. Equation 60 may then be written in the following form

$$C(s) = \frac{1}{1-e^{-T_1 s}}G_1(s) = \frac{\left(1+\sum_{i=1}^{q-1} e^{-\frac{ib}{q}T_2 s}\right)}{1-e^{-bT_2 s}}G_1(s) \tag{63}$$

The z-transform of this with respect to T_2 becomes

$$z_{T_2}[C(s)] = \frac{1}{1-e^{-bT_2 s}} \times \left\{ z_{T_2}\left[\left(1+\sum_{i=1}^{q-1} e^{-\frac{ib}{q}T_2 s}\right)G_1(s)\right]\right\} \tag{64}$$

which may be evaluated by means of tables or the convolution integral.[7] A numerical example of such a system is given in Appendix III. Sampled-data systems with delayed or with more than two samplers may be analyzed by a similar extension of the techniques outlined in the preceding sections.

Conclusions

In the preceding, a method for determining the solution of multiple-sampler systems is given. This is based on the p-transform, developed for the exact solution of linear system with finite pulse width. In the foregoing, open-loop systems are considered only. However, such systems have been treated generally, no restriction being imposed on the number or periodicity of the samplers that they might include. Current investigations have shown that the p-transform may be used to solve linear systems with feedback loop. Results of these will be published in the near future.

In this report it has been shown that:

1. Table III, together with relations 21 and 22, may be used to determine the solution of multiple-sampler systems in which all samplers are periodic, having the same period T. When the samplers have operating delays, corresponding relations of Table IV must be used.

2. Tables III and IV together with relations 21, 22, 49, and 50 may be used to find the solution of multiple-sampler systems in which all the samplers are periodic but have different periods. The final results are expressed in a closed form.

3. Systems with nonperiodic samplers lead to results expressed in infinite series. Such systems become algebraically difficult to handle when they contain more than one nonperiodic sampler.

4. The limiting cases are readily obtained as the pulse widths of the samplers are allowed to approach their limits.

5. Relations 49 and 50 may be used to derive the solution of multiple-sampler sampled-data systems, in which the samplers have different periods, and the pulse widths are assumed to be zero.

Appendix I. Periodic Samplers With Different Periods

Fig. 5 shows a system with two periodic samplers having different periods T_1 and T_2. Let

$$R(s) = \frac{1}{s}$$

$$G_1(s) = \frac{1}{s+1} \tag{65}$$

$h_1 = 0.6$ second
$T_1 = 1$ second
$h_2 = 0.2$ second
$T_2 = 0.4$ second (66)

$R_{p_1}(s)$ is obtained from relation 2 of Table I as

$$R_{p_1}(s) = \frac{1-e^{-h_1 s}}{1-e^{-T_1 s}}\frac{1}{s} \tag{67}$$

Therefore

$$C(s) = \frac{1-e^{-h_1 s}}{1-e^{-T_1 s}}\frac{1}{s(s+1)} \tag{68}$$

$C(s)$ is the input to the second sampler which has a period T_2 different from T_1. Therefore, it is necessary to express equation 68 in terms of T_2 in order to be able to use Tables III and IV. From equation 66 it is apparent that

$$\frac{T}{T_2} = \frac{1}{0.4} = \frac{5}{2}, \text{ i.e., } b = 5 \text{ and } q = 2 \tag{69}$$

Utilizing relation 49, $C(s)$ may be written as

$$C(s) = \frac{(1-e^{-h_1 s})(1+e^{-6/5T_2 s})}{1-e^{-5T_2 s}}\frac{1}{s(s+1)} \tag{70}$$

Multiplying out the numerator, and substituting

$$h_1 = 0.6 = 1.5T_2 \tag{71}$$

equation 70 becomes

$$C(s) = \frac{1}{(1-e^{-5T_2 s})} \times \left(\frac{(1-e^{-4T_2 s})}{s(s+1)} - \frac{(e^{-T_2 s} - e^{-2T_2 s})e^{-0.6T_2 s}}{s(s+1)}\right) \tag{72}$$

With the use of relations 3(c) and 3(a) of Table III respectively, $C_{p_2}(s)$ the output of the second sampler may be written

$$C_{p_2}(s) = \frac{1}{(1-e^{-5T_2 s})} \times \left[(1-e^{-4T_2 s})\,\mathrm{P}_{h_2,T_2}\left(\frac{1}{s(s+1)}\right) - e^{-2T_2 s}(1-e^{-T_2 s})\mathrm{P}_{h_2,T_2}\left(\frac{e^{+0.6T_2 s}}{s(s+1)}\right)\right] \tag{73}$$

The p-transform of the functions in the bracket may be obtained from relations 4 and 5 of Table I. Equation 73 then

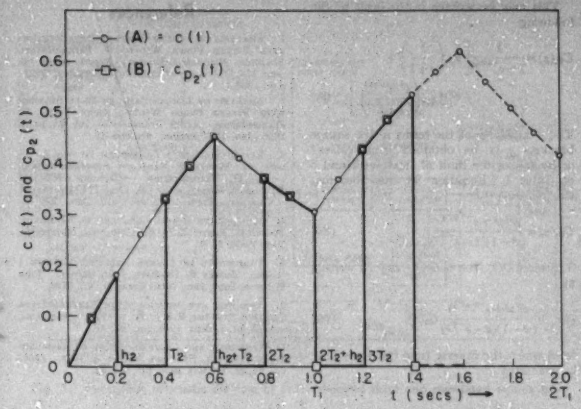


Fig. 6. Curve A—Input to sampler given by equation 68 as function of T_1. Curve B—Corresponding sampler output given by equation 74 as function of T_1

reduces to

$$C_{p_2}(s)=\frac{1}{(1-e^{-sT_2})}\times\left[(1-e^{-sT_2})(1+e^{-sT_2}+e^{-2sT_2}-e^{-3sT_2})\times\frac{(1-e^{-h_2s})}{s(1-e^{-T_2s})}-(1-e^{-T_2s(2h+0.8)})\times(1-e^{-T_2s})-e^{-4sT_2}\times\frac{(1-e^{-h_2(s+1)})}{(s+1)(1-e^{-T_2(s+1)})}\right] \quad (74)$$

Therefore $Q(s)$, and hence, $q(t)$ may be obtained from the relation

$$Q(s)=C_{p_2}(s)G_2(s) \quad (75)$$

It is interesting to check the results obtained from the inverse of equation 68, expressed as a function of T_1, and equation 74 expressed as a function of T_2, for $c(t)$ and $c_{p_2}(t)$ respectively, and to compare them with those physically expected.

Directly from relation 3 of Table II is obtained $c(t)$, the inverse of equation 68, and it may be written as

$$c(t)=[1-d(-0.6)]\times\left[n-\frac{1}{(\epsilon-1)}(\epsilon^{-(m-1)}-\epsilon^{-t})\right] \quad (76)$$

Table V is a sample calculation showing the procedure used for evaluating equation 76. Fig. 6, curve A is the plot of $c(t)$.

Either by the use of the modified z-transform or by the ordinary Laplace inversion integral, $c_{p_2}(t)$, the inverse of equation 74, may be evaluated. For the purpose of this comparison, however, it is sufficient to determine $c_{p_2}(t)$ for a finite interval of time. Under this condition the ordinary Laplace inversion method is simpler and hence more advantageous.

Equation 74 may be expanded, simplified, and written as

$$C_{p_2}(s)=\left(\frac{1-e^{-h_2s}}{s}(1+e^{-T_2s}+e^{-2T_2s}+\ldots)-\frac{1-e^{-h_2(s+1)}}{(s+1)}(1+0.67e^{-T_2s}-0.37e^{-2T_2s}+0.57e^{-3T_2s}\ldots)\right) \quad (77)$$

Thus for the finite interval of time $0\leq t\leq 3T_2$, $c_{p_2}(t)$, the inverse time function may be written

$$c_{p_2}(t)=[1-d(-h_2)][1+d(-T_2)+d(-3T_2)]\times 1-[1-0.82d(-h_2)]\times[1+0.67d(-T_2)-0.37d(-2T_2)+0.57d(-3T_2)]\epsilon^{-t}| \quad (78)$$

Curve B, Fig. 6 shows the plot of $c_{p_2}(t)$, calculated from this equation. It is evident that the relation between $c(t)$ and $c_{p_2}(t)$ satisfies those expected from physical considerations.

Referring to Fig. 5, it should be noted that if a third sampler was present then relations 22, 49, and Table III may be used again to determine $Q_2(s)$. Also, when the samplers have operating delays, corresponding relation of Table IV may be used to find the correct solution.

Appendix II. Example of Nonperiodic Systems

Fig. 7 shows a 2-sampler system. The first sampler is nonperiodic while the second is assumed to be periodic with period T and pulse width h seconds. Let

$$R(s)=\frac{1}{s+\alpha}$$

$$G(s)=\frac{1}{s+\gamma} \quad (79)$$

$R(s)$, the output of the nonperiodic sampler, is obtained from equation 7 of Table I as

$$R_n(s)=\sum_{n=0}^{\infty}\frac{e^{-T_n(s+\alpha)}-e^{-(T_n+h_{n+1})(s+\alpha)}}{(s+\alpha)} \quad (80)$$

where $T_0=0$, since no operating delay is assumed. Therefore $C(s)$ may be written

$$C(s)=\sum_{n=0}^{\infty}\{e^{-T_n(s+\alpha)}-e^{-(T_n+h_{n+1})(s+\alpha)}\}\times\frac{1}{(s+\alpha)(s+\gamma)} \quad (81)$$

This, when expanded, becomes

$$C(s)=\frac{1}{(s+\alpha)(s+\gamma)}-\frac{e^{-h_1(s+\alpha)}}{(s+\alpha)(s+\gamma)}+\frac{e^{-T_1(s+\alpha)}}{(s+\alpha)(s+\gamma)}-\frac{e^{-(T_1+h_2)(s+\alpha)}}{(s+\alpha)(s+\gamma)}+\ldots \quad (82)$$

Equation 82 shows that the input to the sampler is composed of an infinite number of terms, each of which has a prescribed delay. Since it was assumed that the second sampler is periodic, then relation 2 of Table III may be used to determine the sampler output corresponding to each of the input terms. For example, assuming

$$\begin{matrix}h_1<T\\ h_1>h\end{matrix} \quad (83)$$

the first two terms of the sampler output may be written as

$$C_{p_2}(s)=P_{h,T}\left(\frac{1}{(s+\alpha)(s+\gamma)}\right)-e^{-T_1s}\times P_{h,T}\left(\frac{e^{(T-h_1)s}}{(s+\alpha)(s+\gamma)}\right)+\ldots \quad (84)$$

If the second sampler had an operating delay, then the corresponding relations of Table IV could be used to find $C_{p_2}(s)$.

The foregoing illustrates how systems with nonperiodic samplers may be treated by a term-by-term procedure. However, it also illustrates that the equations become complicated if more than one nonperiodic sampler is present for, then, equation 7 must again be used to find $C_{p_2}(s)$, leading to double infinite series.

Appendix III

The limiting case, as h_1 and h_2 are allowed to approach zero, is investigated for the system shown in Fig. 5. Let the system

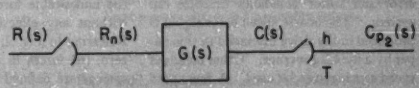


Fig. 7. System having nonperiodic sampler

constants be the same as those given by equations 65 and 66.

$C(s)$, the output of the first stage, may then be written as

$$C(s)=\frac{1}{1-\epsilon^{-T_1 s}}\,\frac{1}{s+1} \tag{85}$$

Relation 49 may now be utilized to express $C(s)$ as a function of T_2.

With the ratio of T_1 to T_2 given as the following

$$\frac{T_1}{T_2}=\frac{5}{2} \tag{86}$$

this becomes

$$C(s)=\frac{1+\epsilon^{-2/2^T s}}{1-\epsilon^{-5/2^T s}}\,\frac{1}{s+1} \tag{87}$$

The output of the second sampler is given by the z-transform of equation 87 taken with respect to the period T_2.

This may be written in the form of the following

$$C_{T_2}(z)=\frac{1}{1-\epsilon^{-5T_2 s}}\left[z_{T_2}\left(\frac{1}{s+1}\right)+\epsilon^{-2T_2 s}\,z_{T_2}\left(\frac{\epsilon^{+0.5T_2 s}}{s+1}\right)\right] \tag{88}$$

The z-transform of the terms in the square brackets may be obtained from tables,[7] or by taking the limit of relations 2 and 3 of Table I. Equation 88 then becomes the following

$$C_{T_2}(z)=\frac{z^2(z^2+\epsilon^{-0.5T_2})}{(z^5-1)(z-\epsilon^{-T_2})}\bigg|_{z=\epsilon^{sT_2}} \tag{89}$$

Therefore $Q(s)$, the output, may be written as

$$Q(s)=\frac{z^2(z^2+\epsilon^{-0.5T_2})}{(z^5-1)(z-\epsilon^{-T_2})}\,G_z(s)\bigg|_{z=\epsilon^{sT_2}} \tag{90}$$

from which the inverse time functions may be evaluated.

References

1. Analysis of Linear Sampled-Data Systems with Finite Pulse Width, G. Farmanfarma. *Electronic Research Laboratory Report*, series 60, issue 157, University of California, Berkeley, Calif., Apr. 1956.
2. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, G. Farmanfarma. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 808–19.
3. Application of p-Transform to the Analysis of Multiple Sampler Systems: Open Loop, G. Farmanfarma. *Electronic Research Laboratory Report*, series 60, issue 174, University of California Press, Berkeley, Calif., Jan. 1957.
4. Multi-Rate Sampled Systems, G. M. Krane. *Technical Report T-14/B*, Columbia University, New York, N. Y.
5. Transients in Linear Systems, Volume I (book), Murray F. Gardner, John Barnes. John Wiley & Sons, Inc., New York, N. Y., 1942.
6. Analysis and Synthesis of Sampled-Data Control Systems, E. I. Jury. *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 332–46.
7. Synthesis and Critical Study of Sampled-Data Control Systems, Eliahu I. Jury. *Ibid.*, vol. 75, pt. II, July 1956, pp. 141–51.

Dynamic Representation of a Hydraulic Constant-Speed Drive

PAUL E. SMITH, JR.
ASSOCIATE MEMBER AIEE

Synopsis: The hydraulic constant-speed drive is an important component of many air-borne electric systems; it is used to convert the varying aircraft engine speed into a constant alternator shaft speed needed to maintain constant system frequency. To study the behavior of electric systems with isolated alternators and other systems with paralleled alternators, it has been necessary to find descriptions of the system components which may be translated into mathematical terms in a formulation convenient for block diagram representation and analog computer simulation. Descriptions of voltage regulators, exciters, alternators, and system load, have been developed and presented in the literature.[1–6] The purpose of this paper is the presentation of information about the hydraulic constant-speed drive and some associated components. The basic drive equations pertain to the ball drive, the axial piston drive, and to the radial piston drive. The equations of the ancillary equipment are pertinent to a particular model.

A FUNCTIONAL schematic drawing of a typical drive and associated equipment is presented as Fig. 1. Speed error and power unbalance activate the system. The coarse speed error signal is produced in the drive discussed in this paper by a flyball governor. Other means of speed sensing may be used. A fine speed error is produced by a frequency-sensitive network and combined with the power unbalance signal and then amplified by a magnetic amplifier. The output of the magnetic amplifier drives a servomotor whose output is added to the coarse error at the governor. The governor output signal is the input to a single-stage hydraulic amplifier whose output varies the stroke of the hydraulic pump part of the transmission.

The governor, hydraulic amplifier, and drive will be described in somewhat more detail in order to facilitate the derivation of their block diagrams. The block diagrams for motor fine-frequency error-sensing network, power-sensing systems, magnetic amplifiers, and servomotor, are relatively straightforward.

The flyball governor is shown in Fig. 2. The upward flyball force F_1 is counteracted by the force exerted by the load and speeder springs and the unwanted forces due to acceleration of moving parts and friction. For the analysis of this system, the undesirable forces may be neglected without any effective loss of accuracy. Referring to Fig. 2, when x, y, and z are zero, the forces in the speeder and load springs are defined as F_{20} and F_{30}. The motion x is taken as zero when the pilot piston of the hydraulic amplifier is centered. By assuming the forces acting on the speeder rod, the relationship of x to y and z is found to be

$$x=\frac{(F_{20}+F_{30}-F_1)+K_2 z+K_3 y}{K_2+K_3} \tag{1}$$

where F_1 is the upward flyball force due to rotation. F_1 is approximately equal to the following

$$F_1=F_{10}(2\omega_a/\omega_{om}-1) \tag{1A}$$

The force F_{10} is the flyball force exerted when the frequency is exactly 400 cps (cycles per second). The quantity ω_{om} is equal to 200π. The alternator shaft speed in radians per second is defined as ω_a. Equation 1(A) is nearly correct so long as the frequency is almost equal to 400 cps, a condition which is met in practice. The other constants are defined in Fig. 2. The load force referred to the fine error servomotor shaft (called the trimmer motor) is negligible in determining the output of the motor over the normal range of operation. The relationship

Paper 57-777, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 4, 1957; made available for printing September 5, 1957.

Paul E. Smith, Jr., was formerly with Massachusetts Institute of Technology, Cambridge, Mass., and is now with the Conval Corporation, Cambridge, Mass.

This work would not have been possible but for the assistance of Westinghouse Electric Corporation and Sundstrand Aviation, who furnished information about alternators and drives, WADC Analog Computer Branch of Aero Research Laboratory, where the simulation was carried out, and the United States Air Force, who sponsored the research program under Contract *AF 33(616)3242*.

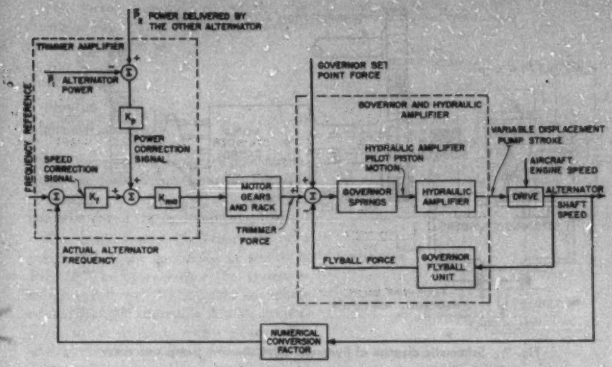


Fig. 1. Functional schematic diagram of constant-speed drive and associated equipment

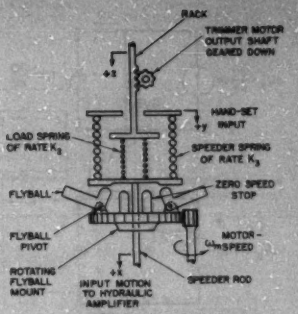


Fig. 2. Schematic diagram of flyball governor

between trimmer motor shaft positions and s is thus given by

$$s = \frac{\theta_{om}}{2\pi N K_{ps}}$$

where K_{ps} is the gear ratio reduction between motor shaft and the shaft of the pinion; N is the number of revolutions required of the pinion to move the rack one inch; and θ_{om} is in radians.

The trimmer motor position θ_{om} is related to motor input by the equation

$$\theta_{om} = \frac{K_m E_m}{s(\tau_m s + 1)} \qquad (2)$$

where

K_m = motor gain, radians per second per volt
τ_m = motor time constant, seconds
E_m = voltage applied to the motor from the magnetic amplifier

The voltage E_m is related to power error in kilowatts and the fine frequency error in cps by the following expression

$$E_m = \left[\frac{K_{ma}}{\tau_{ma} s + 1}\right]\left[K_f(f - f_0) + K_p(\bar{P}_2 - \bar{P}_1)\right] \qquad (3)$$

where

f = actual system frequency, cps
f_0 = desired system frequency, cps = 400
$\bar{P}_1$ = output power of alternator driven by this drive, kw
$\bar{P}_2$ = output power of second alternator connected in parallel
K_f, K_p, and K_{ma} are constants.

The motion x causes a displacement of the hydraulic amplifier pilot piston from the center position. As a result, oil flows to one of the two cylinders whose outputs

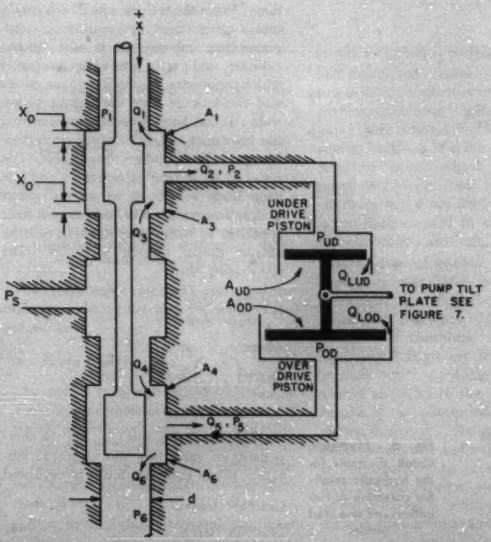


Fig. 3 (left). Schematic diagram of hydraulic amplifier

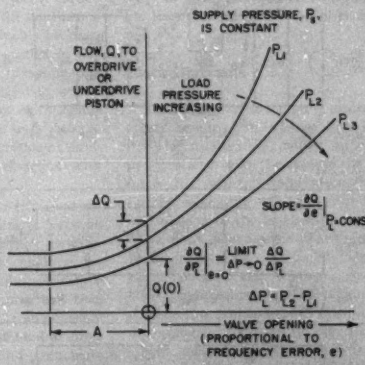

SUPPLY PRESSURE, P_s, IS CONSTANT
FLOW, Q, TO OVERDRIVE OR UNDERDRIVE PISTON
LOAD PRESSURE INCREASING
SLOPE = $\frac{\partial Q}{\partial e}\Big|_{P_L}$ = CONSTANT
LIMIT $\frac{\Delta Q}{\Delta P_L}$ as $\Delta P \to 0$
$\Delta P_L = P_2 - P_{L_1}$
VALVE OPENING (PROPORTIONAL TO FREQUENCY ERROR, e)

Fig. 4. Typical flow versus control valve opening characteristics. Load pressure shown as a parametric variable

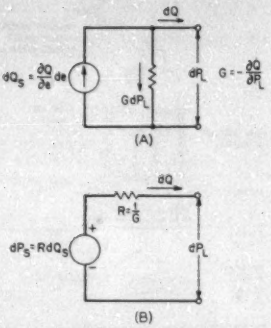


Fig. 5. Representation of one orifice of control valve

A—Equivalent current source
B—Equivalent pressure source

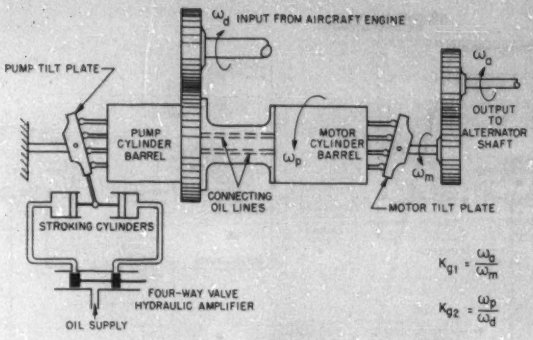


Fig. 7. Schematic diagram of hydraulic drive showing pump and motor

are coupled to the pump stroke so as to change the pump stroke and bring about a correction of the speed of the alternator shaft. The pilot piston and amplifier output pistons are shown in Fig. 3. Flows and pressures are defined in this figure.

It is convenient for the electrical engineer to think of the two orifices comprising the hydraulic valve as two sources with appropriate internal impedance. The parameters of the source are derived from typical flow-opening pressure curves as given in Fig. 4. The flow to the load for an orifice with curves as shown in Fig. 4 can be expressed as

$$Q = C\sqrt{(A+e)^2 + B^2}\,\sqrt{P_s - P_L} \qquad \text{for } -A < e$$

$$Q = CB\sqrt{P_s - P_L} \qquad \text{for } e < -A \quad (4)$$

where A is a constant, proportional to valve underlap, B is a constant proportional to valve radial clearance, and C is a constant proportional to the orifice coefficient.

The representation of the orifice as a source and impedance arises by taking the complete derivative of equation 4

$$dQ = \frac{\partial Q}{\partial e} de + \frac{\partial Q}{\partial P_L} dP_L = dQ_s - G dP_L \qquad (5)$$

where the quantities dQ_s and G are defined in Fig. 5. The small signal or linearized equivalent circuit representation of equation 5 is presented in Fig. 5, together with an equivalent representation in terms of a pressure source.

The flow, delivered by the source of Fig. 5, flows in part into a capacitor representing the compressibility effect in connecting lines; the remainder flows through a resistor representing fluid friction in the line. From the resistor, the flow is partly into a capacitance representing the compressibility of the oil in the output cylinder, and partly into a paralleled load which represents the motion of the stroke and the leakage past the output piston wall. It has been found that connecting line resistance and capacitance are negligible as is the cylinder capacitance. The result of neglecting capacitances and line drop leads to Fig. 6 as an equivalent hydraulic circuit. The ideal transformer is needed to satisfy the equation relating output piston motion and flow. Let $v_p = dx_t/dt$; then

$$v_p = \frac{Q_{OD}}{A_{OD}} = \frac{Q_{UD}}{A_{UD}} \qquad (6)$$

where

$$Q_{OD} = \left(\frac{A_{OD}}{A_{UD}}\right) Q_{UD}$$

x_t = stroke = piston motion

and the flows, Q_{UD} and Q_{OD}, are the components of valve flow due to piston motion. When compressibility is neglected

$$Q_{OD} = Q_3 - Q_{LOD}$$

$$Q_{UD} = Q_3 - Q_{LUD} \qquad (6A)$$

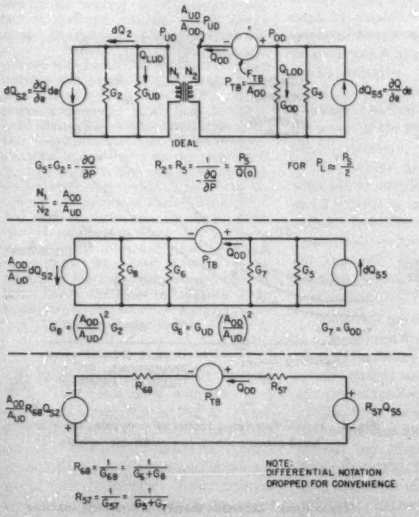


Fig. 6. Hydraulic circuit diagrams for the hydraulic amplifier consisting of two orifices and two load pistons

The force required to move the stroke is F_{TB}. This can be expressed in terms of pressures as

$$F_{TB} = A_{OD} P_{TB} \tag{7}$$

$$P_{TB} = P_{OD} - \left(\frac{A_{UB}}{A_{OD}}\right) P_{UD} \tag{8}$$

By applying successive source transformations, the first circuit shown in Fig. 6 is reduced to the third form which specifies the behavior of the hydraulic amplifier.

The transmission itself consists of a pump and motor connected as shown in Fig. 7. The operation of the transmission is as follows: The pump is capable of delivering a reversible flow q_p proportional to the stroke x_s, which is the output of the hydraulic amplifier. The pump and motor cylinder barrels rotate as a unit, at speed ω_p, driven by the aircraft engine through gearing. The pump delivers an oil flow q_p given by

$$q_p = x_s \omega_p d_p \tag{9}$$

where d_p is the displacement in inches^3 per second per inch of stroke (measured at the stroking cylinder) per radian per second. If the engine speed, and hence pump speed, is at a value different from that required to produce correct alternator speed, the stroke x_s will be nonzero. The resultant flow q_p will cause the motor tilt plate to revolve relative to the cylinder barrel, thus adding or subtracting from the pump cylinder barrel speed an amount sufficient to bring about the correct alternator speed. Part of the pump flow is lost as leakage flow and part as compressibility flow.

The equation which describes this situation is

$$q_p = x_s d_p \omega_p = q_m + \frac{V}{B} sP + G_L P \tag{10}$$

where

V = volume of oil under compression
B = bulk modulus of oil under compression
G_L = leakage coefficient determined from steady-state torque-slip curves

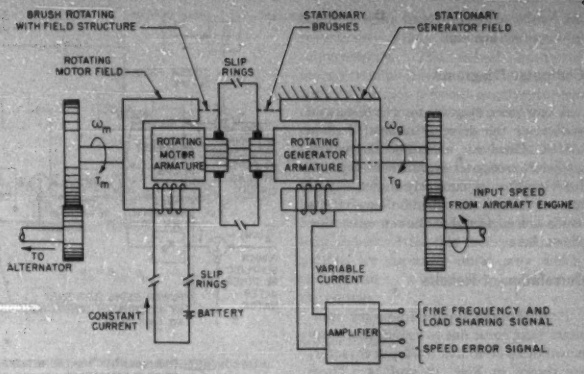
The motor flow expressed in terms of the relative velocity between the parts of the motor is

$$q_m = d_m(\omega_m - \omega_p) \tag{11}$$

The motor torque which equals $d_m P$ is used to overcome: 1. stiction torque which inhibits relative speed and is a nonlinear function of the relative speed, 2. inertia torques; and 3. the electromagnetic torque T_{ω_0} required to turn the alternator shaft.

The drive exhibits several nonlinear effects in addition to the stiction torque just mentioned. Some of the nonlinearities and other effects which may be present but which have been neglected in this paper are: finite stiffness of connecting shafts, backlash of spine coupling, a freewheeling clutch which prevents transmission of power from the alternator to the drive, and a dry-friction type of oscillation damper used in conjunction with the alternator quill shaft.

It is often of interest to electrical engineers to search for analogies between devices in other fields and their own circuits and machines. The analogy between d-c machines and piston-type hydraulic pumps (or motors) has been pointed out.[4] Applying the analogies of Gould and Chu to the differential drive results in the connection of d-c machines shown in Fig. 8. The analog quantities, both parameters and variables, are listed in the figure. The problems of stiction and uneven torque occurring at the straight-through condition in the hydraulic drive seem analogous to the problems of brush friction and cogging that are encountered in low speeds in d-c motors.



ANALOGOUS PARAMETERS		ANALOGOUS VARIABLES	
ELECTRICAL	HYDRAULIC	ELECTRICAL	HYDRAULIC
ARMATURE CIRCUIT TOTAL RESISTANCE R_{mg}	LEAKAGE L	GENERATOR INTERNAL VOLTAGE E_g	PUMP FLOW q_p
ARMATURE CIRCUIT TOTAL INDUCTANCE L_{mg}	VOLUME ÷ BULK MODULUS $\frac{V}{B}$	BACK E.M.F. E_b	MOTOR FLOW q_m
BACK E.M.F. CONSTANT OF MOTOR K_m	DISPLACEMENT d_m	ARMATURE CURRENT I_m	PUMP PRESSURE P_m
GENERATOR CONSTANT VOLTS PER FIELD AMPERES K_g	PER UNIT DISPLACEMENT d_p	GENERATOR SPEED ω_g	ω_p

Fig. 8. D-c machine drive analogous to the hydraulic constant-speed drive

Block Diagrams

The block diagram for the system discussed here is given in literal form in Fig. 9. The number 33, which appears in in Fig. 9(B) is twice the value of the flyball force, F_{fb}, which occurs when the frequency is exactly 400 cps. The linearized version of the block diagram of the drive alone, wherein the effects of clutch, backlash, hydraulic motor stiction, generator friction and windage, generator shaft compliance and damping, and servomotor position and rate limiting are neglected, appears in Fig. 10. The engine speed is assumed constant. The factors leading to negligibly small time constants have been neglected.

Substitution of numerical values into the diagram of Fig. 10 for a drive used with a 40-kva aircraft alternator leads to the result given by Fig. 11, in which

the entire constant-speed drive system is taken into account.

Computer Diagrams

A computer diagram for the linearized version of the drive as simulated on a *GEDA* (Goodyear electronic differential analyzer) computer is shown in Fig. 12. Investigation is currently under way of means of simulating the nonlinear effects which are neglected in the representation shown here.

Correlation of Results

Plots of some typical frequency transients that occur due to application or removal of 20 or 40 kva (0.75 power factor) are shown in Figs. 13 and 14. These curves are selected from many measured responses and are indicative of the correlation between simulated and actual hardware responses. A comparison of simulated and actual transients indicates three significant discrepancies. First, although the computed curves have the proper maximum drop (or rise) in frequency, the time at which the maximum drop occurs is somewhat larger than that indicated in the records of the behavior measured on the actual alternator. Second, the actual alternator response shows a rapid drop (or rise) of short duration (0.01 second) at the beginning of the transient. This effect is not observed in the simulated behavior. Third, the frequency of the actual alternator takes a longer time to return to normal following removal of load than is indicated by the simulated response.

The reasons for these discrepancies may be explained as follows: In addition to the difficulty of making accurate parameter measurements, there is doubtless considerable variation in parameter values from one unit to another. The computer settings were readjusted from the original values which were based on parameter values furnished by the manufacturer. The readjustment procedure was carried out by making successive changes in computer settings until reasonable agreement between simulated and actual response was established. The procedure was then arbitrarily terminated in order to proceed with the simulation of paralleled alternators.

The rapid drop or rise at the beginning of the transients is probably due to a signal introduced into the frequency sensing and recording equipment by the relatively faster transient in line voltage magnitude that occurs when the electrical load is added or removed. The measuring equipment sensed the frequency from the alternator terminal voltage.

The third discrepancy is probably due to neglecting the action of the clutch in the drive representation discussed in this paper. When load is removed, the alternator speed increases, since the drive torque remains initially unchanged due to the energy storage attributable to the compressibility of the oil. To slow down the alternator, negative torque is required from the drive (the drive must accept the energy stored in the alternator inertia). Theoretically, the constant-speed drive, like a Ward-Leonard drive, is a reversible device. The type of simula-

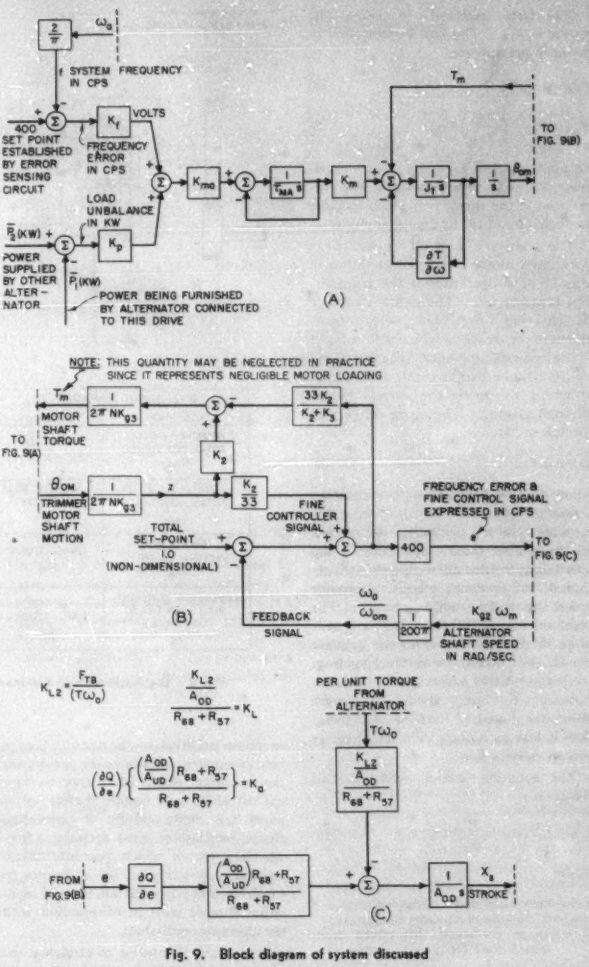


Fig. 9. Block diagram of system discussed

A—Fine frequency error detector, load unbalance circuit, magnetic amplifier, and trimmer motor
B—Hydraulic flyball governor
C—Hydraulic amplifier corresponding to schematic of Fig. 6

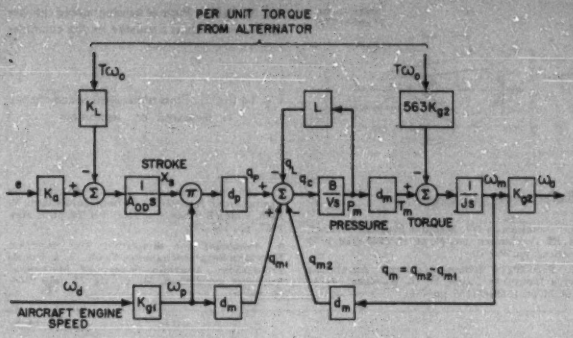


Fig. 10. Linearized block diagram of pump and motor with rigid shaft connection to alternator

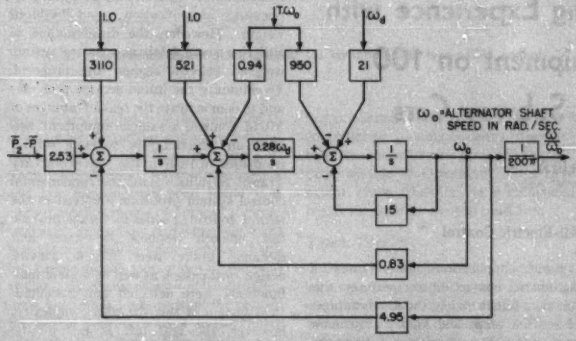


Fig. 11. Block diagram for entire hydraulic constant speed system with numerical values substituted and various simplifying assumptions made

tion discussed in this paper is based on reversibility solely for convenience in representation. The actual hardware incorporates an overrunning clutch of the sprag type which prevents negative torque delivery (or power acceptance) by the drive. Therefore the only retarding torque tending to slow the inertia in the declutched condition is that due to friction and windage, or about 3% to 4% of full-load torque. The nearly linear fall-off in frequency, following the removal of load, substantiates this theory. Investigations are currently under way which should lead to a computer simulation of the clutch.

Conclusions

The simulation studies undertaken to date indicate that satisfactory analog simulation of the drive, alternator, load, and regulators has been achieved. Simulation studies already performed have yielded transients for armature terminal voltage, field current, load current, regulator voltages, etc., for single alternators as well as alternators in parallel. Correlation with results obtained on actual equipment has been as good if not better than that of the frequency transients reported here.

The approach used in the simulation studies carried out so far has been guided by the need to retain the identity of the physical variables and parameters as well as the individuality of the component units making up the whole system even at the expense of increased computer complexity. An advantage of this approach is the ease with which computer "trouble-shooting" may be accomplished, since the open-loop behavior of single units may be

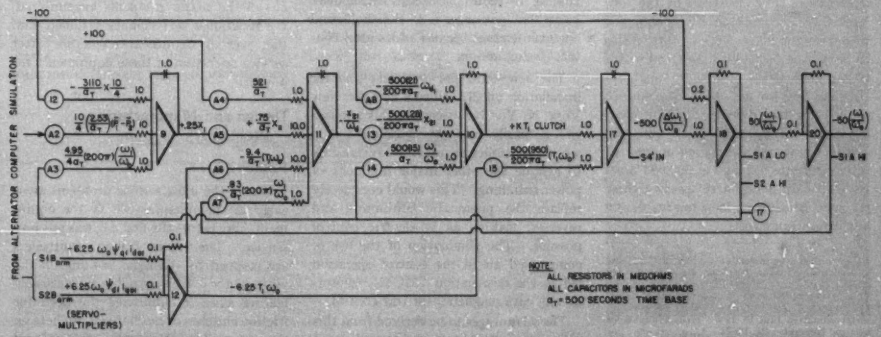


Fig. 12. Computer diagram for linearized constant-speed drive for use on Goodyear electronic differential analyzer computer

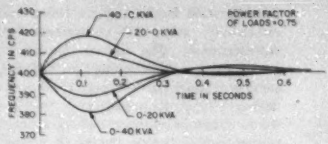


Fig. 13 (left). Plots of transient speed changes for various loads as simulated on the computer

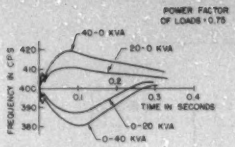


Fig. 14 (right). Plots of transient speed change as measured on equipment

studied independently of the remaining parts of the system. Furthermore, this approach allows for a ready substitution of different components into the simulated system without revising the entire computer setup.

References

1. "Dynamics of D-C Machine Systems, M. Riaz. *AIEE Transactions*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 365–70.

2. Full-Wave Bridge Magnetic Amplifiers with Inductive Loads, H. H. Woodson. *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 7–9.

3. Analog Computer Representation of Alternators for Parallel Operations, Luiz V. Boffi, Vinton B. Haas, Jr. *Ibid.*, vol. 76, pt. I, May 1957, pp. 153–58.

4. Analogies for Hydraulic and Electric Drives in Servomechanisms, Y. Chu, L. A. Gould. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 75, July 1953, pp. 851–57.

First Year's Operating Experience with New Control Equipment on 100 New York City Subway Cars

G. W. WEBER
ASSOCIATE MEMBER AIEE

THE OPERATION of multiple-unit electric trains in modern high-speed rapid-transit service demands an extremely high degree of equipment reliability. This performance must be achieved with a minimum of time and labor for maintenance. Efforts to achieve these objectives have, in recent years, produced notable advances in the design of both propulsion and control equipment.

The type of car currently being placed in service in large quantities by the New York City Transit Authority is equipped with four 100-horsepower motors. This amount of power requires a control with a resistance switching mechanism capable of handling 250 kw and providing smooth accelerating and dynamic braking characteristics. This mechanism must be able to operate through a cycle of approximately 30 steps as many as 500,000 times per year. In addition to meeting these requirements, it must have low inspection and maintenance costs.

Paper 58-250, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 17, 1957; made available for printing November 21, 1957.

G. W. Weber is with the General Electric Company, Erie, Pa.

All-Electric Control

Earlier applications used banks of pneumatic contactors to perform this function, but these had the disadvantages of too few steps and high maintenance costs. In 1947 a pilot-motor-driven cam-operated resistance-stepping controller was developed to meet these requirements. This radically new, discrete notching controller provided simultaneously sufficient notches to reduce acceleration impulses below the "passenger feel" threshold, and low maintenance because of its ideal contact closing action.

The new controller was first sold for installation on 375 subway cars for New York, N. Y. Its success in this application naturally led to the proposal that a second controller be added on future cars to assume the other major functions of power switching. This would completely replace the pneumatic contactors and reverser, making an all-electric control possible. The elimination of the use of compressed air in the control operation led to the designation "MCM" (motor-driven, cam magnetic) for this control.

The advantages to be derived from this new system offered powerful incentives for its adoption. They included decreased maintenance, standardization of components, simplification, and reduced weight. However, the disadvantage of adopting a completely new control system weighed heavily against the proposal. To minimize the initial service problems and to demonstrate the real advantages of MCM control, a sample equipment was built and operated in revenue service from 1952 to 1955 on the New York City Transit System. Here the fundamental control system problems involved in the remote control of series motors to provide fast, smooth response in power and dynamic brake were solved. Device design weaknesses, as well as system malfunctions, were detected and corrected. In a system of this size and complexity, however, the real test is not in the sampling but in the full-scale use of it. Thus, in the manufacture and initial servicing of the first 100 MCM equipments for the New York subway many problems developed. This paper discusses the major problems encountered, the functional performance attained, and the over-all maintenance picture after service operation of these equipments for a year.

Design and Manufacturing Problems

Factory Test

One of the most serious problems arose early in the manufacture of the equipment, long before the first car was put into service. The motor-driven controllers are stopped by dynamic braking on the pilot motor. This is a highly desirable feature since it eliminates the need for friction clutches or mechanical ratchets on the controller. However, this leaves only the inherent friction of the camshaft

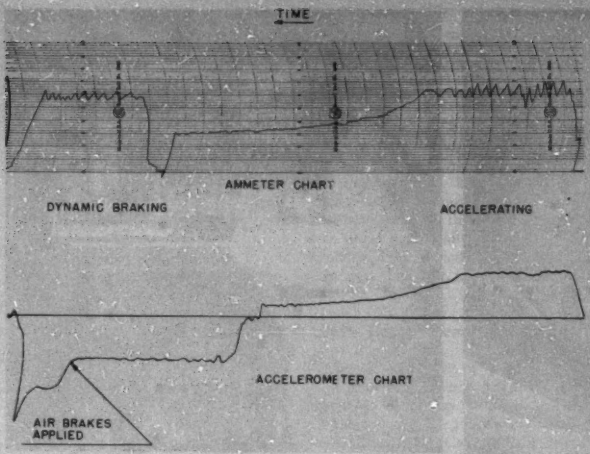


Fig. 1. Performance charts of rapid-transit car equipped with type-MCM control

to resist any tendency of the controller to creep under the influence of vibration and the loading reaction from the cam contours. Conditions of adverse cam loading and low friction were soon found that caused notch skipping on some of the controllers. A sweeping redesign of the cams was made to reduce peak camshaft loading torque in the stop zones of the controller. Since this change, 600 controllers have been built and operated for periods up to one year without a single known case of notch skipping from this cause.

Each control group was given a rigorous functional test by means of special test equipment designed to reproduce the car circuits and operations. In addition to its primary function of detecting manufacturing defects, this test equipment also proved valuable in detecting design weaknesses. In several instances control equipments failed to respond positively to test conditions, revealing loopholes in the controller positioning circuitry, where intolerance manufacturing variations could cause system malfunctions. To a large extent, these difficulties were eliminated before the first car went to track test.

Track Test

During the track testing of the first ten cars, operating conditions were found that could cause the new magnetic line circuit breakers to pump under load. The circuit at fault was designed to reduce line-breaker coil-excitation current and thus minimize load on the control train line wires. A new "optimized" line-breaker coil design permitted the elimination of this circuit and of the problem it had caused.

The first cars were processed through track test almost without incident. Control response in both accelerating and dynamic braking proved to be fast and smooth. Typical ammeter and accelerometer charts are shown in Fig. 1. These smooth accelerating and braking characteristics rendered interaction between cars practically nonexistent. Track test adjustments were quickly and easily made by means of the small trackside door, Fig. 2, which gave easy access to equipment for all adjustments. As many as four cars per day could be processed by the same track test crew if necessary.

Service Experience

To date, 100 cars equipped with this control have been in service on the Brooklyn-Manhattan Division of the New York City Transit Authority for approximately 18 months. An additional 200 cars have been in service on the Interboro Rapid Transit Division for approximately 9 months. These cars are operating an average of 70,000 miles per car per year in extremely severe service.

With 300 equipments on what amounts to a mass life test, even the slightest design weakness or malfunction quickly becomes apparent. For example, shortly after the cars entered revenue service, operators found that the line breakers were tripping occasionally at third-rail gaps. Investigation revealed that a third-rail gap occurring at a critical point in the acceleration cycle could cause the controllers to go out of position. During

Table I. Unattended Service Life of MCM Equipment

Device	Part	Quantity	Unattended Service Life, Miles
Line circuit breaker	Contacts	4	50,000*
	Arc chute	2	100,000
	Arc horns	2	100,000
Controller: main circuit switches	Contacts	4	100,000
Controller: notch interlocks	Contacts	4	100,000
Current limit relay	Contacts	2	100,000*
Magnetic contactor	Contacts	2	100,000
Relays	Contacts	10	100,000
All other contacts and wearing parts			over 500,000

* Estimated.

Fig. 2. Trackside door conveniently located to facilitate making necessary adjustments

Fig. 3. Control equipment mounted underneath a rapid-transit car, covers removed to show accessibility of equipment

Fig. 4. View underneath a car on the inspection pit, showing accessibility of control equipment and convenient working position for maintainer

the reapplication of power after passing the third-rail gap, two traction motor armatures could be connected directly across the line without their fields, which resulted in a flashover. Once recognized, the problem was completely eliminated by a simple interlock change.

As would be expected, there have been a number of difficulties of the "growing pains" variety. These have not been more numerous or serious than might be expected when a radically new equipment of such magnitude is placed in so exacting a service. Some, like the afore-mentioned have been easily corrected as soon as recognized. Others have required further investigation and occasional changes in details of equipment design or circuitry.

Whenever a trouble has come to light, it has been dealt with in a manner that will preclude further occurrence. While a few difficulties have not been completely cleared up as yet, intensive effort is being devoted to them with good prospects of early correction.

Maintenance

Present Picture

Recently an evaluation was made of the maintenance requirements of the MCM control equipment. Each part that is subject to wear or erosion was studied and its "unattended life" evaluated. The results of this analysis are shown in Table I.

A bogey of 100,000 miles has been set for the minimum maintenance period or unattended life of every control component. Only the line breaker has fallen short of this to date. At present its minimum is estimated at 50,000 miles. Further improvements in this area are being studied, but the extreme duty on these breakers—approximately ten load interruptions per mile—may make the 100,000-mile goal impractical.

The established inspection period on equipment of previous types was 3,000 miles. As a result of recent maintenance study, an inspection period of 7,500 miles has been adopted for the new equipment, and a 15,000-mile period is being evaluated. Inspection periods as high as 30,000 miles may ultimately be found to be practical.

Of particular interest is the unattended life of the large number of resistance controller contacts. Because of the non-arcing system design and the ideal contact

Fig. 5. New York City Transit Authority subway car, showing type-MCM control equipment installed under floor

closing mechanics, these contacts will last well over 500,000 miles. Such contact life is not considered possible in devices that pound on closing or arc on opening; or where partial closure can cause spitting, overheating, or welding. With the discrete notching controller, the camshaft travel is divided into alternate "run" or "switching" zones and "stop" zones where all contacts are fully closed. The controller cannot hesitate or stop in a "run" zone with contacts partially closed.

Inherent Advantages

The long maintenance-free life of this control is due in part to the elimination of the internal wearing parts inherent in pneumatic devices. The critical wearing parts of magnetic relays and contactors, and of motor-operated controllers, are readily accessible for inspection and maintenance; see Figs. 3 and 4. Therefore, frequent inspections for the purpose of detecting hidden troubles before they cause failures are unnecessary. The maintenance man can replace visibly worn parts and close the MCM control with full confidence that no hidden troubles will grow into failures before the next inspection.

The oil can, another big maintenance headache, had been eliminated. This is of particular advantage where the air is heavily laden with brakeshoe dust. Periodic cleaning is made infinitely easier by the absence of oil drippings and splatter. To date, no routine lubrication of the control equipment has been found necessary.

Trouble-shooting on this control has been made easy through the use of a small test unit that can be slipped onto the control under the car. With this unit, a sequence test can be made in a few minutes that is so complete that it will easily detect 99 per cent of all propulsion control troubles. This test unit has proved invaluable in dealing with initial service problems.

Future Developments

Control equipment of the MCM type is readily adaptable to all-electric car operation. This has been demonstrated by the two sample cars now in operation on the lines of the Chicago (Ill.) Transit Authority. These cars utilize dynamic brakes (80 to 2 mph) as the primary brake, eliminating air-brake equipment entirely. If this scheme were adopted on large rapid-transit cars of the type used in the New York subway, a major reduction in weight and maintenance costs would be achieved by elimination of the air-brake equipment. The principal obstacle at present is the need for a reliable high-capacity, electrically operated brake for holding and emergency duty, and to supplement the dynamic service brake if necessary.

Conclusions

A survey of the service history of the first 100 MCM control equipments placed in service indicates that they have proved capable of successful performance under the severest operating conditions. Acceleration and braking are remarkably smooth, and control response is fast.

A maintenance study on these same equipments indicates that original objectives are being achieved. There is every indication that added experience will permit further extension in inspection intervals, thus giving fuller realization of the advantages inherent in this type of equipment.

In addition to the 300 equipments presently in service, 125 more are on order and specifications are being prepared for several hundred additional cars.

References

1. A New Control for Heavyweight Rapid-Transit Cars, I. W. Lichtenfels. *AIEE Transactions*, vol. 72, pt. II, May 1953, pp. 49–53.

2. High-Speed Rapid Transit Equipment, S. J. Vonch. *Ibid.*, vol. 73, pt. II, Nov. 1954, pp. 267–70.

Discussion

R. K. Gerlitz (General Electric Company, Erie, Pa.): In the 15 months since this paper was first presented, at the 1956 AIEE Summer and Pacific General Meeting in San Francisco, Calif., an additional 125 MCM equipments have been placed in service, and 225 MCM equipments are presently being installed. These 350 equipments are identical to the previous 300 equipments with the exception of the method of providing battery charging during car lay-up. The previous 300 cars provided a constant 4-ampere charging rate into the 37.5-volt batteries by means of the 600-volt third-rail system and a voltage-dropping resistor. This gave a 2,250-watt loss in the dropping or current-limiting resistor. The new cars utilize the motor-generator set to provide battery charging with a result that losses are now limited to normal motor-generator set losses. In addition, charging rate during lay-up is no longer limited to 4 amperes, but is determined by battery requirements.

The intervening 15 months have also furnished additional data concerning the "unattended life" of the limiting maintenance items. In the development of the MCM control system, a fundamental concept was to confine all possible load-current interruptions to the two magnetic line circuit breakers. Normal inspection and maintenance of contact tips would then be limited to the line breakers, rather than a procedure involving replacement of all power contact tips and arc chutes. The one exception to this concept in practice, on the MCM equipment, is the full series transition contactor.

Table II displays the degree to which this has been accomplished.

Table II. Unattended Service Life for MCM Equipment

Device	Part	Quantity	Unattended Service Life, Miles
Line circuit breakers	contacts	4	50,000
	arc chute	2	200,000
	arc horns	2	300,000
Controller; main circuit switches	contacts	4	100,000
Controller; notch interlocks	contacts	4	100,000
Current limit relay	contacts	2	125,000
Magnetic contactor	contacts	2	150,000
Relays	contacts	10	150,000
All other contacts and wearing parts			over 500,000

G. W. Weber: Mr. Gerlitz's discussion of this paper reveals that the "unattended service life" of the critical parts of the MCM control system has been materially bettered. This progress is largely the result of continued development and across-the-board improvements in the design of the basic components of the system. Although the line breaker contacts remain the limiting service item, it is significant to note that this item of equipment is separately mounted and is easily accessible for inspection and maintenance from either track-side or maintenance pit.

Limiting Values of Driving-Point Impedances and Transfer Functions Due to Component Variations

SHELDON JONES
ASSOCIATE MEMBER AIEE

Synopsis: A method is suggested which determines the limiting values of driving-point impedances and transfer functions due to component variations. The method consists of a procedure for obtaining a set of equations, by a conformal mapping technique, which defines the limiting cases of component tolerances.

ONE of the most acute problems in building complex, precision systems is in obtaining parts and components that can be manufactured in quantity with only small deviations about an established norm. These components, in general, must perform as specified under a variety of adverse conditions. Generating a set of specifications for individual or groups of components is seldom an easy task. The effect of tolerances on over-all system performance must be maintained within allowable limits. However, overspecifying is not the answer to the tolerance problem. This practice is wasteful and creates superfluous procurement problems.

In designing electric circuits it is often important to know the bounds on the magnitude and phase of impedance networks; for example, it is necessary to know impedance variations when designing devices to drive these networks. A method is suggested in this paper which establishes such bounds.

The effect of variations on gain and phase of transfer functions is particularly important in closed-loop servo systems for stability considerations. The effect of component tolerances and frequency variations on the magnitude and phase of transfer functions can be determined precisely by the method proposed. Conversely, allowable tolerances can be set on these quantities.

The method presented here consists of finding a set of equations that defines the bounds on the impedance or transfer function due to circuit tolerances. These equations are determined by a conformal mapping technique using the limiting values of the components as established by the tolerances. This technique always results in a set of equations which defines the limiting cases; each individual equation describes a line or a circle, and these lines or circles form a closed region showing the variation of the impedance or transfer function.

Two-Terminal Networks

The method is illustrated by determining the bounds on the input impedance of Fig. 1(A). Assume that R, L, and C may vary independently and may assume any value between the limits defined in equation 1.

$$R_1 \leq R \leq R_2 \tag{1A}$$

$$L_1 \leq L \leq L_2 \tag{1B}$$

$$C_1 \leq C \leq C_2 \tag{1C}$$

The impedance variation of the RL branch of Fig. 1(A) is restricted to the crosshatched area shown in Fig. 1(B). A $1/Z$ transformation of the crosshatched (impedance) area yields the admittance bounds (the image[1] of the impedance bounds) of the RL branch. The input admittance of the circuit in Fig. 1(A) for a specific sinusoidal frequency is

$$\frac{1}{Z} = \frac{j}{X_C} + \frac{1}{R+jX_L} \tag{2}$$

The admittance of the RL branch as defined in equation 2 can be redefined in terms of two new variables u and v.

Let

$$u+jv = \frac{1}{R+jX_L} \tag{3}$$

Solution of equation 3 yields the following four equations

$$u = \frac{R}{R^2+X_L^2} \tag{4}$$

$$v = \frac{-X_L}{R^2+X_L^2} \tag{5}$$

$$R = \frac{u}{u^2+v^2} \tag{6}$$

$$X_L = \frac{-v}{u^2+v^2} \tag{7}$$

Substitution of equation 6 into the lower limit of the inequality equation 1(A) produces

$$v^2+\left(u-\frac{1}{2R_1}\right)^2 \leq \left(\frac{1}{2R_1}\right)^2 \tag{8}$$

This is the equation of a circle in the limiting case with its origin at $v=0$ and $u=1/2R_1$, as is shown in Fig. 2(A). The area enclosed by the circle is the image of the area in the impedance plane for $R>R_1$. Of course the portion of the circle $v>0$ (the dotted portion) corresponds to a negative inductance and is therefore physically unrealizable. Equation 8 gives one of the boundary conditions on the admittance of the RL branch. This corresponds to the boundary condition $R=R_1$ in the impedance plane.

Similarly, the upper limit of the inequality equation 1(A) is

$$v^2+\left(u-\frac{1}{2R_2}\right)^2 \leq \left(\frac{1}{2R_2}\right)^2 \tag{9}$$

This is the equation of a circle in the limiting case with its origin at $v=0$ and $u=1/2R_2$ as is shown in the plot of Fig. 2(B). The area exterior to the circle is the image of the area in the admittance plane for $R<R_2$. The dotted portion of the curve $v>0$ again corresponds to a negative value of inductance and is physically unrealizable.

Similarly, equation 1(B) gives these two equations:

$$\left(v+\frac{1}{2X_{L1}}\right)^2+u^2 \leq \left(\frac{1}{2X_{L1}}\right)^2 \tag{10}$$

$$\left(v+\frac{1}{2X_{L2}}\right)^2+u^2 \geq \left(\frac{1}{2X_{L2}}\right)^2 \tag{11}$$

Equations 10 and 11 are plotted in Fig. 2(C). The area common to the two circles corresponds to the area in the impedance plane defined by equation 1(B). The area bound by the two circles for $u<0$ corresponds to negative resistance and is therefore physically unrealizable.

The crosshatched area bounded by the four circles is plotted in Fig. 2(D). This area represents the maximum deviation of the admittance of the RL branch.

To determine the boundary on the input admittances, the admittance of the capacitance branch is added to the equations defining the RL branch. Since the

Paper 58-81, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted August 13, 1957; made available for printing November 8, 1957.

SHELDON JONES is with the Hughes Aircraft Company, Culver City, Calif.

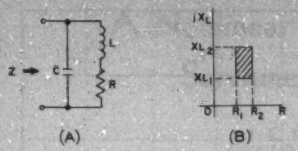


Fig. 1 (left). Driving-point impedance example

A—Parallel tuned circuit
B—Impedance variation of RL branch

admittance is purely imaginary, addition of the term j/X_c represents a translation of the real axis.

When the capacitance assumes its limiting values, eight new equations result. However for normal tolerances on these components (the range of tolerances in examples 1 and 2) the following four equations represent the bounds on the admittance.

$$u^2+\left(v-\frac{1}{X_{C1}}+\frac{1}{2X_{L1}}\right)^2\leq\left(\frac{1}{2X_{L1}}\right)^2 \qquad (12)$$

$$u^2+\left(v-\frac{1}{X_{C1}}+\frac{1}{2X_{L2}}\right)^2\geq\left(\frac{1}{2X_{L2}}\right)^2 \qquad (13)$$

$$\left(u-\frac{1}{2R_1}\right)^2+\left(v-\frac{1}{X_{C1}}\right)^2\leq\left(\frac{1}{2R_1}\right)^2 \qquad (14)$$

$$\left(u-\frac{1}{2R_2}\right)^2+\left(v-\frac{1}{X_{C2}}\right)^2\geq\left(\frac{1}{2R_2}\right)^2 \qquad (15)$$

Transformation back to the impedance plane ($1/Z$) yields the following four equations, where x and y represent the real and imaginary components, respectively, of the impedance, Z

$$x^2+\left(y+\frac{X_{C2}^2-2X_{L2}X_{C2}}{2X_{C2}-2X_{L2}}\right)^2\leq\left(\frac{X_{C2}^2}{2X_{C2}-2X_{L2}}\right)^2 \qquad (16)$$

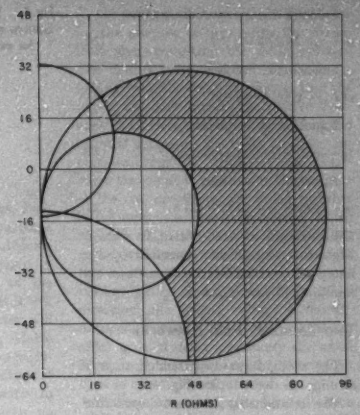


Fig. 3 (right). Impedance variation for example 1

where

$X_{C2}>X_{L2}$

$$x^2+\left(y+\frac{X_{C1}^2-2X_{L1}X_{C1}}{2X_{C1}-2X_{L1}}\right)^2\geq\left(\frac{X_{C1}^2}{2X_{C1}-2X_{L1}}\right)^2 \qquad (17)$$

where

$X_{C1}>X_{L1}$

$$\left(x-\frac{X_{C1}^2}{2R_1}\right)^2+(y+X_{C1})^2\leq\left(\frac{X_{C1}^2}{2R_1}\right)^2 \qquad (18)$$

$$\left(x-\frac{X_{C2}^2}{2R_2}\right)^2+(y+X_{C2})^2\geq\left(\frac{X_{C2}^2}{2R_2}\right)^2 \qquad (19)$$

It is important to note that if the inductive reactance is greater than the capacitive reactance in equations 16 or 17, the inequality sign of that equation must be reversed. In developing the equations, the inequalities were divided by X_C minus X_L. If X_L is greater than X_C, this is a negative quantity and the inequality is reversed. The case where X_L equals X_C is left for the reader. The result, in the limiting case, is an equation for a line. The effects of frequency variations can be evaluated by considering the resulting impedance variation of the frequency dependent elements.

Example 1: Equations 16, 17, 18, and 19 are evaluated to determine the bounds on the driving point impedance of Fig. 1(A). Assume the following conditions

$f=1{,}600$ cycles per second

$10\leq X_L\leq 15$

$13.1\leq X_C\leq 14.4$

$2.3\leq R\leq 3.5$

The above numbers represent a 20% variation in the resistance and inductance and 5% variation in capacitance.

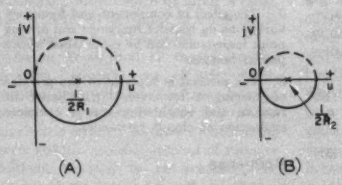
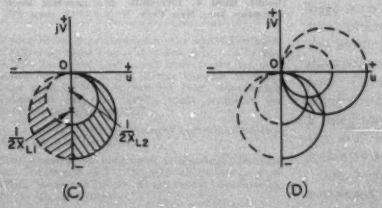


Fig. 2 (left). Conformal map of Fig. 1 (B)

A—Plot of equation 8
B—Plot of equation 9
C—Plot of equations 10 and 11
D—Composite plot



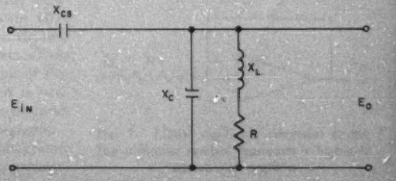


Fig. 4. Two-terminal pair network

Substitution in equation 16 gives

$$x^2+(y+58.2)^2 \geq (45.2)^2 \quad (20)$$

Equation 17 becomes

$$x^2+(y-9.16)^2 \geq (23.6)^2 \quad (21)$$

Equation 18 becomes

$$(x-45.2)^2+(y+14.4)^2 \leq (45.2)^2 \quad (22)$$

Equation 19 becomes

$$(x-24.6)^2+(y+13.1)^2 \geq (24.6)^2 \quad (23)$$

These equations are plotted in Fig. 3. The crosshatched area bounded by the four circles establishes the bounds on the input impedance variation.

Two-Terminal Pair Networks

The method can be extended to a 4-terminal network. See Fig. 4.

The relationship between input and output, where Z is the impedance of the parallel branch, is

$$\frac{E_{IN}}{E_O}=\frac{-jX_{C3}+Z}{Z}=\frac{-jX_{C3}}{Z}+1 \quad (24)$$

Equations 12, 13, 14, and 15 describe the bounds on $1/Z$. Multiplication by the vector j gives a 90-degree co-ordinate rotation. Multiplication by X_{C3} simply alters the magnitude of the co-ordinate. The addition of unity to the equation is a translation of the imaginary axis.

These equations represent the impedance bounds for small variations in components. When the series capacitor is allowed to assume its maximum and minimum values, the equations 25 to 28 describe the bounds on the ratio of input to output.

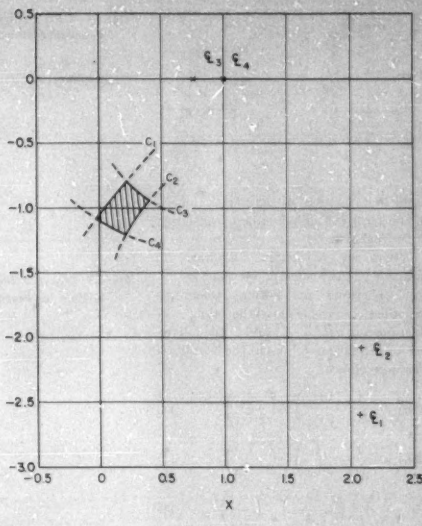
$$y^2+\left[x-1+\frac{X_{C3}(X_{C1}-2X_{L1})}{2X_{L1}X_{C1}}\right]^2 \leq \left(\frac{X_{C3}}{2X_{L1}}\right)^2 \quad (25)$$

$$y^2+\left[x-1+\frac{X_{C3}(X_{C1}-2X_{L2})}{2X_{L2}X_{C2}}\right]^2 \geq \left(\frac{X_{C3}}{2X_{L2}}\right)^2 \quad (26)$$

$$\left(y+\frac{X_{C3}}{2R_1}\right)^2+\left(x-1-\frac{X_{C3}}{X_{C1}}\right)^2 \leq \left(\frac{X_{C3}}{2R_1}\right)^2 \quad (27)$$

$$\left(y+\frac{X_{C3}}{2R_2}\right)^2+\left(x-1-\frac{X_{C3}}{X_{C2}}\right)^2 \geq \left(\frac{X_{C3}}{2R_2}\right)^2 \quad (28)$$

Fig. 3. Variation in input to output ratio for example 2



Example 2: Assume that the circuit of Fig. 4 is to give the transfer function at 1,600 cps as represented by equation 29.

$$E_{IN}=E_O/80^\circ \quad (29)$$

The components are assumed to have a range of values as follows

$$21.8 \leq R \leq 25.18$$

$$43.1 \leq X_L \leq 47.2$$

$$95.2 \leq X_C \leq 105.3$$

$$103.8 \leq X_{C3} \leq 114.6$$

The foregoing numbers represent a 7% variation in the resistance, a 5% variation in inductance and a 5% variation in capacitance. Substitution of the appropriate numbers in equation 25 gives

$$y^2+(x-0.759)^2 \leq (1.33)^2 \quad (30)$$

Similarly, equation 26 gives

$$y^2+(x-0.99)^2 \geq (1.1)^2 \quad (31)$$

and equation 27 becomes

$$(y+2.63)^2+(x-2.09)^2 \leq (2.63)^2 \quad (32)$$

as likewise equation 28 yields

$$(y+2.06)^2+(x-2.09)^2 \geq (2.06)^2 \quad (33)$$

The four equations are plotted in Fig. 5. The variation in the transfer function is limited to the crosshatched area.

Conclusions

Very little analytical effort has been devoted to the unglamorous task of defining and evaluating the effect of component tolerances. It is believed that two valid statements can be made concerning the proposed method:

1. The effect of component and frequency variations on transfer functions and driving point impedances can be established within precise bounds.

2. The method is superior to the practice of varying a few circuit parameters at random and considering this a complete evaluation of circuit tolerances.

Reference

1. Complex Variables and Applications (book). Ruel V. Churchill. McGraw-Hill Book Company, Inc., New York, N. Y., 1948, chap. IV.

A Nonlinear Integrator for Servomechanisms

J. C. CLEGG
ASSOCIATE MEMBER AIEE

INTEGRATION in the forward part of a servomechanism loop has long been known to reduce steady-state errors. With one perfect integrator, there will be no steady-state error following a simple step-function input; with two tandem integrators there will be no steady-state error due to a ramp input, etc. The major drawback to the linear integrator is the time delay involved. Each linear integrator introduces 90 degrees of phase lag at all frequencies, and so it takes only two integrators to make a basically unstable system. A nonlinear integrator is to be described which is superior in this respect to a linear type.

Theory

For purposes of this discussion, an integrator may be considered to have two time delays. One is the time required for the output to reach a certain usable value after a certain input signal is applied. This delay results in attenuation of high frequencies and ordinarily improves stability. The second delay is the time required to bring the output to zero after the input goes to zero. This delay ordinarily serves no useful purpose and is entirely destabilizing. With a perfect linear integrator, the second type of delay is infinite (since a true integrator will hold a constant output when its input is zero), and a reversal of input is required to cause the output to diminish.

As an illustration of the integration problem, consider an electric servomechanism consisting of a linear integrator driven by an error voltage and driving an electric motor. Assume a small initial error and a noticeable amount of static or coulomb friction to be present. The output of the integrator will steadily increase until it reaches a voltage high enough to cause the motor to overcome the friction and begin to turn. Assume the error then to be quickly brought to zero. At this point the output of the integrator will stop increasing, but it will not yet decrease, and so the motor will continue to turn, causing an unavoidable overshoot (and, in this simple system, instability). It is clear that if the output of the integrator could be reset instantly to zero whenever the error goes to zero, no prolonged overshoot would be necessary, and the system stability would be greatly improved. This can be done without sacrificing any of the desirable qualities of integration.

For simplicity, an electronic integrator of the type used in analog computers will be considered. Fig. 1 shows a simple electronic integrator circuit described by the following equation. (The derivation of this equation can be found in any recent elementary text on electronic analog computers.)

$$e_o = \frac{-1}{RC}\int e_i dt \qquad (1)$$

Fig. 2 shows the same circuit with the addition of two diodes. For positive e_i, the diode across the input resistor will not conduct; and because of the polarity reversal of the amplifier, the output voltage will be negative, so the diode across the capacitor will not conduct. Since neither diode conducts, both can be neglected, and equation 1 still holds for Fig. 2 (as long as e_i is positive), and the circuit of Fig. 2 is simply an integrator. As soon as e_i goes through zero and starts to become negative, the diode across R conducts, short-circuiting R. This makes R in equation 1 approach zero and causes the gain of the integrator to approach infinity. With such high gain, it takes but an instant of integration of the slightly negative input voltage to integrate away all of the voltage previously present at the output. The diode across C prevents e_o from going positive. Thus, the circuit of Fig. 2 integrates whenever the input is positive and resets its output to zero almost instantly when the input goes negative.

Fig. 3 shows two nonlinear integrators connected together. The two resistors R_d are very small compared to R and are for the purpose of preventing the diodes from short-circuiting the input voltage. The diodes are connected with different polarities for the two amplifiers so that the upper amplifier integrates positive voltages and resets quickly to zero when the input is negative; the lower amplifier integrates negative voltages and resets when the input is positive. The outputs of the two amplifiers are added together in the summing network consisting of two resistors R_s, and the two amplifiers together will handle either positive or negative input voltages. Whenever the input voltage passes through zero from either direction, the output voltage is quickly dropped to zero. Thus, the first (desirable) type of delay previously discussed is retained, and the second (undesirable) type is effectively eliminated.

The nonlinear integrator has only been seriously considered and tested in its electronic form. However, it seems that almost any integrating device (electric, hydraulic, mechanical, etc.) can be modified to incorporate the ideas presented here by the judicious use of diodes, check valves, 1-way clutches, or rachets, etc.

Discussion

It is shown in the appendix that the describing function of the circuit of Fig. 3 is

$$N = \frac{-1.62}{\omega RC} \angle -38.1 \text{ degrees} \qquad (2)$$

Thus, the nonlinear integrator has only 38.1 degrees of phase lag compared to 90 degrees for a linear integrator; an improvement in phase margin of 51.9 degrees at all frequencies and all amplitudes. With the use of this integrator and a d-c motor, a type *2* servomechanism could be built which has the low-frequency asymptote shown in the Nyquist plot of Fig. 4. Such a system could be stabilized by gain adjustment alone, even though other time lags may be present at high frequencies. Contrast this to the linear type *2* system having the same steady-state performance but whose low-frequency asymptote extends out along

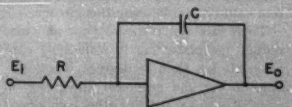


Fig. 1. Linear electronic integrator circuit. The triangular symbol represents a high-gain operational amplifier

Paper 58-10, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted July 22, 1957; made available for printing October 18, 1957.

J. C. CLEGG is with the Ramo-Wooldridge Corporation, Los Angeles, Calif.

This paper is an outgrowth of Research Contract no. *DA-36-039 SC-70113* with the University of Utah, sponsored by the Evans Signal Laboratories, U. S. Signal Corps.

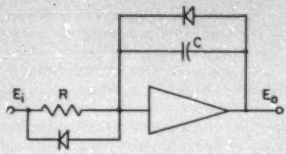


Fig. 2. Nonlinear electronic integrator for positive input voltages

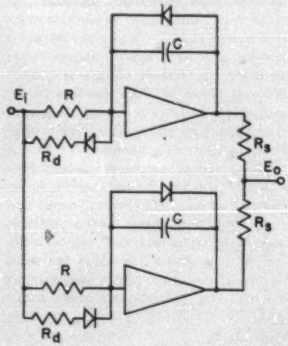


Fig. 3. Nonlinear electronic integrator for positive or negative input voltages

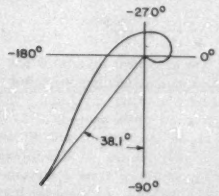


Fig. 4. Polar plot of a possible transfer function of a system using the nonlinear integrator of Fig. 3

the −180-degree axis making it unconditionally unstable without some form of phase lead compensation. It is also conceivable that a basically stable type 4 system could be built using this circuit.

To reduce the response time and also reduce the phase lag to something considerably less than 38 degrees, it is advisable to use the nonlinear integrator in parallel with a resistor or other component which involves no integration. (The polarity reversal in the integrator must be taken into consideration.) The result would be a "proportional plus integral" control.

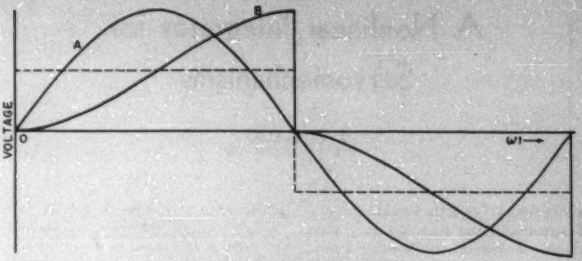


Fig. 5. Waveforms for the circuit of Fig. 3

A—Input sine wave B—Output wave

Test Results

A simple nonlinear proportional plus integral controller using silicon diodes and 1-tube Miller integrator circuits, in place of operational amplifiers, was tested on a simple servomechanism with a high friction load. The integrator was able to reduce the steady-state error by a factor of 10 while introducing only 10 degrees of phase lag at the natural frequency of oscillation of the system.

When the friction load was removed to observe stability, it was found that the nonlinear integrator caused the step-response overshoot to increase from about 29 per cent to only about 37 per cent. When it was attempted to achieve the same reduction of steady-state error by gain increase alone or by use of a linear integrator, the system became unstable as might be expected.

Appendix

In order to calculate the describing function of the nonlinear integrator, assume that the circuit of Fig. 3 works in the ideal manner. That is, assume the output to be the time integral of the input, but to be reset instantly to zero every time the input goes through zero. Neglecting for the present the scale factor $-1/RC$, the waveshapes of input and output will be as shown in Fig. 5 for a unit sinusoidal input.

Over the first half-cycle, the output voltage is

$$E_o = \int \sin \omega t \, dt = -\frac{1}{\omega} \cos \omega t + C_1 \qquad (3)$$

This must be zero when $t=0$, so C_1 must be $1/\omega$. Over the second half-cycle, equation 3 must still hold, requiring the constant term to be $-1/\omega$.

An examination of Fig. 5 shows that the output wave is the sum of a square wave in phase with the input and a cosine wave (a sine wave lagging the input by 90 degrees). The amplitude of the square wave is one half the vertical portion of the output wave.

$$E_{sq} = \frac{1}{2}\frac{1}{\omega}(1-\cos\pi) = \frac{1}{\omega} \qquad (4)$$

The amplitude of the cosine wave is the same $(1/\omega)$.

The fundamental component of a square wave is $4/\pi$ times its amplitude, and of course the cosine wave has only a fundamental component, there being no higher harmonics. The fundamental component of the output voltage is therefore the complex sum of two voltages*

$$E_1 = \frac{4}{\pi}\frac{1}{\omega}\underline{/0 \text{ degrees}} + \frac{1}{\omega}\underline{/-90 \text{ degrees}} \qquad (5)$$

$$= \frac{1}{\omega}\left[\frac{4}{\pi} - j1\right] \qquad (6)$$

$$= \frac{1}{\omega}\sqrt{\left(\frac{4}{\pi}\right)^2 + 1^2}\,\underline{/\tan^{-1} - \pi/4} \qquad (7)$$

$$= \frac{1.62}{\omega}\underline{/-38.1 \text{ degrees}} \qquad (8)$$

Equation 8 also gives the describing function since the input had unit amplitude and zero phase angle.

Reinserting into equation 8 the previously suppressed scale factor, $-1/RC$, results in equation 2.

$$\frac{E_1}{1/0 \text{ degrees}} = N = \frac{-1.62}{\omega RC}\underline{/-38.1 \text{ degrees}} \qquad (2)$$

* The second and all even harmonics are zero. The odd harmonics are simply the harmonics of a square wave: $E_3 = \frac{4}{3\pi\omega}$, $E_5 = \frac{4}{5\pi\omega}$, etc.

Heat Pump and Heating Cables Installed in the Same Residence: First Year of Comparative Operation

C. W. JONES
MEMBER AIEE

E. E. LINDEN
MEMBER AIEE

THIS PAPER is presented upon completion of the first year of alternate operation of a heat pump and heating cables installed in the same residence,[1] located within 10 miles of Providence, Rhode Island; see Fig. 1. Both systems are full capacity and only one was in operation at any time. Operation metered for comparison was started on October 6, 1956 and ended on October 6, 1957, another year of records is underway at the present time. This paper is presented to list the findings of the first year and invite observation or criticism from those with experience within this field in other geographical areas.

At the start of this year, metering facilities had been provided, as discussed in reference 1, in order to contribute data with several objectives in mind:

1. An equivalence in kilowatt-hours between the heat pump on heating and the radiant ceilings (performance factor).

2. A determination of heat losses based on kilowatt demand.

3. The determination of heat-pump coefficient of performance from metered energy input in comparison to manufacturer's specifications and the relationship between heat-pump coefficient of performance and annual performance factor.

4. An indication of total maximum residential kilowatt demand and kilowatt-hour use for load factor computations; and the cost of operation.

5. Observations in connection with the mechanical and electrical operation of the system.

6. Observations in connection with comfort experience of each system.

Paper 58-64, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 18, 1957; made available for printing November 7, 1957.

C. W. JONES and E. E. LINDEN are with The Narragansett Electric Company, Providence, R. I.

The author wishes to acknowledge the work of Brnest Deluski and Lawrence Urban in connection with data compilation and analysis, particularly with the correlation of kilowatt-hour use to weather.

Equivalent Kilowatt-Hours: Heat Pump Versus Radiant Ceilings

In attempting to determine the ratio of equivalence between the heat pump and radiant ceilings, each system was operated alternately with the other throughout the heating season. In this connection the operating schedule in Table I developed.

The two heating plants installed in the same residence presents an equal heating requirement on each plant for the same conditions of weather. It remained to determine what effect various weather factors would have on this residence and it was found that outdoor air temperature was dominant. This is consistent with the conclusions covered in other studies and is used for the computation of equivalence between the two systems.

Outdoor temperature was recorded on a chart drawing instrument located adjacent to the residence. These charts were used as a basis in determining the heating load, in terms of degree-hours below 70 F (degrees Fahrenheit) for the alternate periods of operation. The performance factor between the heat-pump installation and the radiant ceilings was computed by comparing the kilowatt-hours per degree-hour (kw-hr/deg-hr) for one system against the other. Although the duct heaters were operated at times with the heat pump, the kilowatt-hours used by the duct heaters was divided by 2.1 (performance factor developed in the following), in order to represent the combined heat pump and duct heater use as equivalent to all heat-pump operation:

Heat-Pump Operation

Kw-hr use:

Heat pump	5,819
Well pump	1,194
Duct-heater (÷2.1)	110
Other (÷2.1)	248
	7,371 kw-hr

Corresponding degree-hours:
From 7-day charts at the installation 113,802 deg-hr

Kilowatt-hours per degree-hour:

$$\frac{7{,}371 \text{ kw-hr}}{113{,}802 \text{ deg-hr}} = 0.0648 \text{ kw-hr/deg-hr}$$

Radiant Ceiling Operation

Kilowatt-hour use	8,220
Corresponding degree-hours	59,996

Kilowatt-hours per degree-hour:

$$\frac{8{,}220 \text{ kw-hr}}{59{,}996 \text{ deg-hr}} = 0.137 \text{ kw-hr/deg-hr}$$

Comparison

$$\frac{\text{Radiant ceiling kw-hr/deg-hr}}{\text{Heat-pump kw-hr/deg-hr}} = \frac{0.137 \text{ kw-hr/deg-hr}}{0.0648 \text{ kw-hr/deg-hr}} = 2.11$$

All meters were read on a weekly basis and a plot was made for each heating period of kilowatt-hours versus degree-hours. Fig. 2 shows this plot and a line which represents a linear correlation of these data arrived at through the least-square method. It does not appear to be significantly apparent that the relationship among the data is anything other than linear. Fig. 3 is a temperature duration curve for the period of this report and is offered in comparison with a similar curve presented in reference 1. Table II is a tabular form of Fig. 3.

Fig. 4 is a monthly bar graph showing the average daily degree-hours and the corresponding average kilowatt-hours per degree-hour. In this figure, the periods when the radiant ceilings were in service have been converted to equivalent heat-pump kilowatt-hours by the performance factor of 2.1. It is interesting to note that the kilowatt-hours per degree-hour is fairly uniform from month to month, rising from a low value of 0.050 in the fall, to 0.072 in the midwinter, and falling again to 0.065 in the late spring. It is believed that the difference between the spring and fall values (when the weather is about the same) could be related to either or both of the following factors:

1. In the fall the heat pump was in use and controlled by an indoor thermostat whose differential between "on" and "off" was found to be as much as 4 degrees. Under this condition the heat pump was on for longer periods of time than when controlled by the present thermostat (in service in mid-November) with a differential of about 1½ degrees. Similarly when controlled by a close differential thermostat the number of starts is increased. It follows, then, that the fall operation probably required fewer kilowatt-hours per degree hour due to the wide thermostat differential, since it takes a heat pump a period of time to come up to its optimum coefficient of performance on each start.

2. While the heat conducted through the walls and ceilings of the residence is probably the same per degree temperature difference in the fall as in the spring, there is some reason to believe that the radiation losses of the outer surfaces of the house might

Fig. 1. Residence where heating equipment is installed

be higher in the spring, since the ground and other adjacent bodies, such as trees, are colder then than in the fall, due to a thermal lag from season to season.

Heat Losses Computed from Data

The development of a heat loss value from data is derived from the least-square curve shown in Fig. 2. The following equation relates heat-pump kilowatt-hours of use to degree-hours of experience for periods which average one week:

$$\text{Kw-hr/week} = 0.07068 \times \text{degree-hours/week} - 29.44$$

For a one-hour period this equation becomes:

$$\frac{\text{Kw-hr/week}}{168 \text{ hours/week}} = \frac{0.07068 \times \text{degree-hours/week} - 29.44 \text{ kw-hr/week}}{168 \text{ hour/week}}$$

For a 70 F temperature difference this become (0 F design):

$$\frac{\text{kw-hr}}{\text{hr}} = (0.07068 \times 70 - 0.175)$$

$$\frac{\text{kw-hr}}{\text{hr}} \text{ (heat pump)} = 4.77$$

$$\frac{\text{kw-hr}}{\text{hr}} \text{ (heat pump)} \times \text{performance factor} = \text{heat loss}$$

$$4.77 \text{ kw-hr/hr} \times 2.1 = 10.05 \text{ kw-hr/hr heat loss}$$

$$= 34{,}200 \text{ BTU/hr heat loss}$$

Selecting a safety factor of 20% from the dispersion of points in Fig. 2 the heat loss to be matched by heating equipment becomes:

$$1.20 \times 34{,}200 \text{ BTU/hr} = 41{,}000 \text{ BTU/hr heat loss}$$

Reference 1 lists calculated heat losses for the residence and these are referred to here for comparison:

A nearby electrical utility . . . 42,000 BTU/hr
Heating cable manufacturer . . . 48,000 BTU/hr
Reliable heating contractor . . . 75,000 BTU/hr

In explaining these loss figures it is interesting to note that the "utility" calculation was based on actual experience with electrically heated residence and that the contractor calculated 57,000 BTU/hr but would definitely not recommend anything under 75,000 BTU/hr.

Heat-Pump Coefficient of Performance

The heat pump under consideration is a nominal 3-ton machine. Its published heating specification is 42,000 BTU per hour on 9 gpm (gallons per minute) of 50 F water. The condensing unit is rated at 3 hp (horsepower), however the kilowatt power demand for the total machine is not listed. The rating is further based on a specified 1,200 cfm (cubic feet per minute) of circulating air.

After installation was completed and specified conditions of water and air established, the coefficient of performance was checked several times in January 1957 under the instrumentation shown in reference 1. The greatest temperature drop in the water recorded was 5 F. Fig. 5 shows the location of immersed-bulb thermometers. The heat pump including circulating air fan and control relays drew a demand of 3.70 kw; and the well pump operated at 0.77 kw. Considering first the COP (coefficient of performance) without the well pump and assuming that the electric energy delivered to the heat pump (excluding the well pump) is completely converted to heat:

BTU per hour from water:

$$9 \text{ gpm} \times 60 \text{ min/hr} \times 8.33 \text{ lb/gal} \times 5 \text{ F} = 22{,}500 \text{ BTU/hr}$$

BTU per hour electric:

$$3.70 \text{ kw-hr/hr} \times 3413 \text{ BTU/kw-hr} = 12{,}600 \text{ BTU/hr}$$

Total BTU per hour input

$$= 35{,}100 \text{ BTU/hr}$$

The COP (excluding well pump)

$$= \frac{\text{Input: electric plus water}}{\text{Input: electric}}$$

$$= \frac{35{,}100 \text{ BTU/hour}}{12{,}600 \text{ BTU/hour}}$$

$$= 2.79$$

Now if the water pump is included in the electric input but not included in the heat input:

The COP (including well pump)

$$= \frac{35{,}100 \text{ BTU/hour}}{15{,}200 \text{ BTU/hour}}$$

$$= 2.31$$

It is of further interest to develop

Table I. Operating Schedule of Heat Pump and Radiant Ceilings

Period	Heat Pump, Days	Radiant Ceilings, Days
Oct. 6 to Dec. 22	76	
Dec. 22 to Jan. 5		14
Jan. 5 to Jan. 20	15	
Jan. 20 to Jan. 26		6
Jan. 26 to Feb. 24	29	
Feb. 24 to Mar. 10		14
Mar. 10 to Apr. 2	23	
Apr. 2 to June 2		61
Totals	143	95

Table II. Temperature Duration: Oaklawn, Rhode Island, October 1956–September 1957

Temperature, F	Number of Hours Below	% Heating Load Below
100	8,760	
90	8,726	
80	8,402	
70	7,847	
65	6,594	100
60	5,824	88.5
50	4,406	67
40	3,025	46.9
35	2,283	34.7
30	1,433	21.8
25	777	11.8
20	457	6.9
15	253	3.8
10	148	2.2
5	77	1.2
0	32	0.5
−5	13	0.2
−10	2	0.1

the COP based on published rating sheets representing factory tests on the compressor alone. Representative head and suction pressures are: 160 psi (pounds per square inch) and 31 psi

Refrigeration capacity	26,300	BTU/hr
Power input (3.07 kw)	10,500	BTU/hr
Total	36,800	BTU/hr

$$\text{The COP (compressor only)} = \frac{\text{Power input plus refrigeration capacity}}{\text{Power input}}$$

$$= \frac{36{,}800 \text{ BTU/hr}}{10{,}500 \text{ BTU/hr}}$$

$$= 3.50$$

The contents of the 2.31 COP is the same as the contents used in developing the performance factor of 2.1 between the heat pump and radiant ceilings.

In September 1957 the air-fan speed was increased to deliver 1,860 cfm and the output improved. However, since the duct work was not designed to handle this amount of air, a noise problem developed. The manufacturer is increasing the surface area in the condenser and evaporator in order to reduce the air volume to 1,200 cfm and provide 42,000 BTU per hour. As soon as this new unit is tested at the factory it will be substituted for the present unit.

Kilowatt Demand and Load Factor

Due to the trial nature of the first year of operation and difficulties with the chart-drawing demand-meter during the period of coldest weather, figures shown for kilowatt demand and load factor are of necessity estimates. Kilowatt-hour use records are without this difficulty; however, the equivalence factor of 2.1 is used between heat pump and radiant ceilings in order to show a full year comparison of one to the other.

Kilowatt-Hour Use and Costs 1956–57: Table III

With heat pump operating alone:		
Domestic	12,799 kw-hr	$281.59
Heat-pump	13,732 kw-hr	$274.64
Total	26,531	$556.23
With radiant ceilings operating alone:		
Domestic	12,799 kw-hr	$281.59
Radiant-ceilings	25,007 kw-hr	$500.14
Total	37,806	$781.73

Maximum 15-Minute Kilowatt Demand and Load Factor

Heat-pump operation		
Maximum 15-minute demand	12	kw
Corresponding load factor	0.25	
Radiant-ceiling operation		
Maximum 15-minute demand	16	kw
Corresponding load factor	0.27	

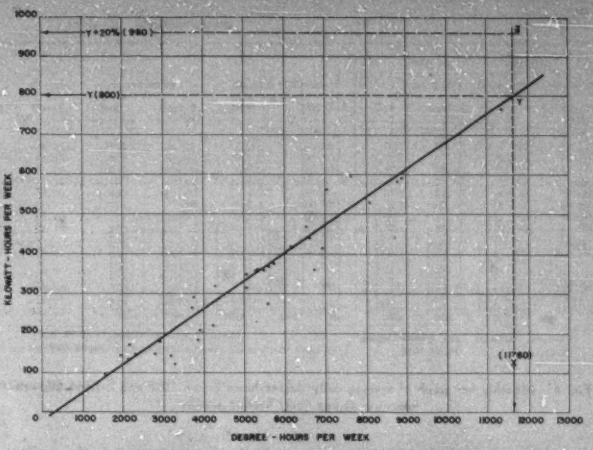


Fig. 2. Scatter diagram and best linear curve for kilowatt-hour use versus degree-hours below 70 F, weekly periods. Kw-hr/week = 0.07069 deg-hr/week, 29.44

Equipment Operation

Several remarks are in order in connection with the various equipments involved in the two heating systems, beyond that no major troubles were encountered.

HEAT PUMP

1. Relays and controls were satisfactory for completely automatic operation.

2. The heating thermostat was changed in mid-November 1956 due to an unpleasant operating differential of 4 degrees. The new thermostat was supplied with heat anticipators providing a differential of under one degree. This caused an unnecessarily large number of heat-pump starts and stops, however after removing the "anticipator" element a differential of a degree-and-one-half was produced and considered satisfactory.

3. Noise from the compressor found its way upstairs through the return air ducts. The installation of a one-inch blanket of fuzzy insulation inside the return air duct for a distance of 5 feet back from the

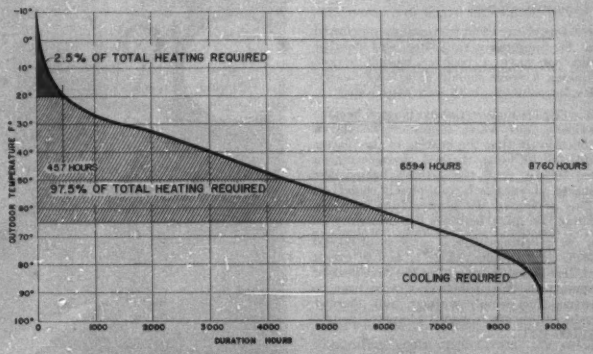


Fig. 3. Temperature duration curve October 1956 to October 1957, one year

X—One week averaging O F outdoors
Y—800 kw-hr use (period X)
Y+20% allows for deviation from mean curve
Z—Safe heating capability for O F design

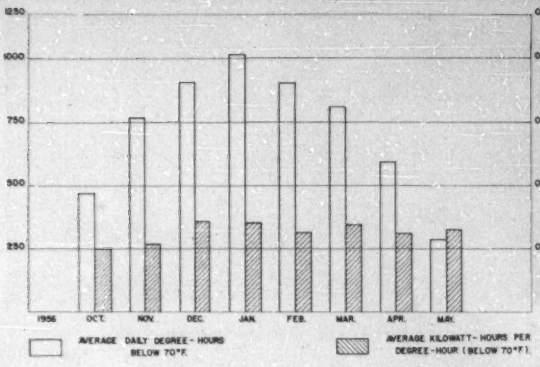


Fig. 4. Monthly bar graph of average daily degree-hours below 70 F and average kilowatt-hours per degree hour, heating months

machine should absorb the noise.

4. Plans are being made to test the power requirements of a submersible-type well pump in comparison with the jet-type presently in service. It appears from manufacturer's published ratings that a submersible-type well pump might reduce the pumping costs as much as 50% or about $15 per year.

5. The addition of turning vanes in the air duct elbows closest to the heat-pump resulted in a smoother, increased flow, permitting a lower fan speed, and thereby reducing fan and air noise.

6. In place of the two steps of duct-heater (1–3 kw step and 1–4 kw step) mentioned in reference 1, one step of duct-heater, 4 kw at 240 volts, is found adequate and set to come on at an outdoor temperature of 0 F. This assures adequate heating capacity at all times.

Radiant-Ceilings

1. One of the room thermostats was found to be improperly calibrated and another did not have a positive closing action, making a noise similar to a faint buzzer. Replacements for both were satisfactory.

2. Cracks in the ceiling plaster became evident, and where it cannot be shown conclusively that the heating of the plaster was not at fault, there is stronger reason to believe that the lathers butted the sheets of rock-lath too closely together. This prevented the plaster from properly filling the space between the sheets.

3. The ceilings have a capacity of 60,000 BTU at 240 volts. This is well in excess of heat-losses indicated by the first year of operation. A transformer will be used to reduce the supply voltage and thereby reduce the ceiling capacity to approximately 41,000 BTU.

Comparative Cost of Heating

For the heating season from October 1956 through May 1957 the total kilowatt-hour requirement of the radiant ceiling plant was as follows:

24,007 kw-hr: Cost $480.14

Reflecting a 2.1 performance factor between the heat pump and radiant ceilings, the corresponding requirement was:

11,432 kw-hr: Cost $228.64

In comparison, the local gas rates are quoted as $5.64 for the first cubic feet and $1.19 per 1,000 cubic feet thereafter monthly, 1,050 BTU per cubic foot. The minimum would be exceeded in each of the months under consideration and an efficiency of 80% is assumed:

8 mo of minimum	16,000 cu ft	$ 45.12
Remainder	81,700 cu ft	97.20
Total gas cost		$142.32

Fuel oil averages 15.5 cents per gallon during the period. A heating value of 140,000 BTU per gallon and 70% efficiency is assumed:

837 gallons of oil	$129.80
$2.00 per month burner power	16.00
Total oil cost	$145.80

Comfort experience with both systems was good as reported by the family of two adults and three children living there.

Heat-Pump Comfort

From a temperature-humidity recorder it was found that the humidity indoors seldom fell below 30 per cent relative, even during cold spells. The master heating thermostat carried a 72 F setting except on days when the temperature fell below zero and then the setting was moved up to 74 F.[1]

A problem developed in connection with air temperature stratification where the downstairs living room (in the basement) is directly connected to the upstairs family dining room kitchen through a wide open stair well. In this case a thermostat was installed downstairs to control the air-circulating fan. After a heating cycle, as stratification takes

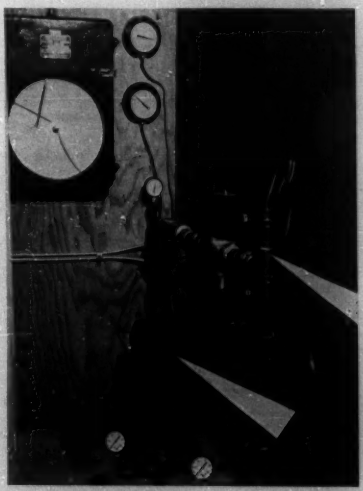


Fig. 5. Location of immersed-bulb thermometers for determining the temperature differential between supply and discharge water

Table III. Operating Costs

Heating Period	Use	Domestic† Black‡	Domestic† Red¶	Cost,$	Heat Pump Use	Heat Pump Cost, $	Radiant Ceilings Use	Radiant Ceilings Cost, $	First Year Operating Costs If on Heat Pump Use	If on Heat Pump Cost, $	If on Radiant Ceilings Use	If on Radiant Ceilings Cost, $
Oct. 1956	1,023	589	434	22.22	744	14.88	1,560	31.20	1,767	37.10	2,583	53.42
Nov. 1956	1,050	630	420	23.35	1,200	24.00	2,520	50.40	2,250	47.35	3,570	73.75
Dec. 1956	1,116	682	434	24.56	2,045	40.90	4,300	86.00	3,161	65.46	5,416	110.56
Jan. 1957	1,023	589	434	22.22	2,420	48.40	5,060	101.60	3,443	70.62	6,103	123.82
Feb. 1957	896	504	392	19.70	1,680	33.60	3,530	70.60	2,576	53.30	4,426	90.30
Mar. 1957	899	465	434	19.12	1,674	33.48	3,515	70.30	2,573	52.60	4,414	89.42
Apr. 1957	930	510	420	20.10	1,110	22.20	2,330	46.60	2,040	42.30	3,260	66.70
May 1957	992	558	434	21.44	559	11.18	1,172	23.44	1,551	33.62	2,164	44.88
Total heating	7,929	4,527	3,402	172.71	11,432	228.64	24,007	480.14	19,361	401.35	31,936	652.85
Nonheating Period* June, July, Aug., Sept. 1957	4,870	3,162	1,708	108.88	2,300*	46.00*	1,000‡	20.00‡	7,170	154.88	5,870	128.88
Total first year	12,799	7,689	5,110	281.59	13,732	274.64	25,007	500.14	26,531	556.23	37,806	781.73

Note: "Use" is in kilowatt-hours.
* Months when heating, cooling, dehumidification, and air fan circulation all occurred but are not metered separately.
† Two dial meter.
‡ Domestic rate total service 7 A.M. to 11 P.M.
¶ 1 cent per kw-hr total service 11 P.M. to 7 A.M.: estimated at 14 kw-hr per day.
‡ For attic fan and some heating, estimated.

place, the downstairs thermostat is set to start recirculating air before discomfort is experienced.

During the summer months the cooling and dehumidification were used as necessary. On pleasant days windows were opened and the machine shut down. Dehumidification was called upon more often than cooling.

Radiant-Ceiling Comfort

After the end of the first week of operation the family had set the various room thermostats to what they considered satisfactory zoning. In general, hallways and bedrooms were set cooler than living areas, lavettes, and the bathroom. The family realizes, however, that this system provides no cooling or dehumidification in the summer. Perhaps an attic fan would prove an incomplete substitute but certainly desirable. The operating costs for the radiant ceiling has been adjusted to include the use of such a circulator, in an attempt to reflect a fair annual comparison.

Operating Schedule for Next Year

New Heat Pump and Reduced Capacity Ceiling

In August 1957 there seemed to be certain wisdom in surveying the manufacturers of heat pumps to see what might be available to replace the unit we had, since substantial data, in connection with the operation of this first unit, had been gathered. After considering the replies to the survey and discussing the availability of an improved model with the manufacturer of the first unit, the decision was made to substitute the improved model of the first unit as soon as it became available. It should be said further that a large proportion of the manufacturers surveyed could provide units with comparable specifications. Ceilings will be operated at a reduced capacity of 41,000 BTU per hour.

Additional Data

Since the first year of heating operation involved the heat pump for almost twice as many hour as the ceilings, the second year will be heavier on radiant-ceiling operation. In fact, consideration is being given to using the radiant ceilings for the entire next heating season. An additional chart drawing demand meter will be installed to meter the radiant ceilings alone, and a 2-register kilowatt-hour meter will also be installed to compare day and night use. Records kept through the next year will be essentially the same as for the first year with the addition of data indicated by the added instrumentation.

Conclusions

Costs

From a cost standpoint, and without laboring the details to the nearest dollar, several generalizations can be made:

Heat-Pump Costs

1. Installation of a heat-pump costs about the same as a combination fuel fired and cooling system.

2. In this instance the operating cost for the heat pump is computed to be 60% higher than oil or gas, with a performance factor of 2.1 for the heat pump.

Radiant Ceiling Costs

3. Installation of radiant ceilings (and perhaps other direct-conversion systems) is less expensive than any other type of fully zoned system.

4. Although the operating costs are roughly twice as much as that experienced in heat pump operation this past year, reduced investment and no apparent need for maintenance or replacement could substantially reduce the annual costs of such a system.

Performance Factor

It appears that the difference between the COP and the performance factor of a heat-pump installation is a function of the number of starts and this in turn is a function of the control thermostat differential. The smaller this differential, the lower will be the performance factor relative to the COP.

Heat Loss Calculations and Load Factor

Calculated heat losses can be very different from actual heat losses in a residence and the calculated figures were all found to be higher in this case. The calculated annual operating costs therefore would be correspondingly higher than actual. However, in the heat-pump case, the performance factor was assumed to equal the COP in the cost calculation and this effected a reduction in calculated operating costs approaching the actual cost figure for the heat pump. In the case of the radiant ceilings, however, there is no compensation for inaccuracies in loss calculation through the difference between

a COP and the performance factor. The result of operating cost calculations for a radiant-ceiling installation has an error directly proportional to the error between calculated and actual losses. This error could be discouragingly large in comparison with alternative heating systems.

The load factor of an electric heating system is directly related to proper sizing. An oversized heating plant may not require many more kilowatt-hours to provide comfort, but its kilowatt demand on the utility's distribution system represents a needless burden, indicating a rate consideration. After more operating experience is gained with this residence, and perhaps others, it may be possible to develop a more realistic formula for calculating local heat losses as well as the safety factor to assure optimum comfort and minimum kilowatt demand.

Heat-Pump Equipment

While, generally speaking, the equipment on the market today is reliable and satisfactory from a comfort standpoint, the cost of the heating operation is still not comparable to fossil fuels in this, and perhaps most, areas. Assuming that utility rates are in line with the kind of service required for heating in any geographical area, there are perhaps two other considerations that the equipment manufacturer could pursue:

1. The development of new refrigerant gasses more suited to the heating task and with higher COPs than the present Freons.

2. The development of improved compressors and auxiliary equipment which would narrow the gap between Carnot Cycle COPs of 6 "and up," to the 2.1 performance factor experienced in the first year operation of the system covered in this paper.

Reference

1. Heat Pump and Heating Cables Installed in the Same Residence for Data Purposes, C. W. Jones, E. E. Linden. *AIEE Transactions*, vol. 76, pt. II, Nov. 1957, pp. 307–14.

Characteristics of Phase-Controlled Bridge Rectifiers with D-C Shunt Motor Load

R. W. PFAFF
MEMBER AIEE

IN INDUSTRY in general, and highly automated industries in particular, there are many applications which require the use of variable-speed electric motor drives. The combination of a phase-controlled rectifier with one or more d-c shunt motors provides a drive which is flexible, highly responsive, and low in weight and cost.

Recently the use of single-phase bridge (diametric, double-way) phase-controlled rectifiers has been gaining favor in the low-horsepower range. When compared to the single-phase diametric full-wave rectifier circuit, the bridge rectifier offers the advantages of elimination of anode transformers and of the necessity for providing separate sources of armature and shunt field power. Elimination of these two items makes considerable cost and weight savings possible.

In the following, a detailed description of the operation of the bridge rectifier with a d-c shunt motor load is presented. Also, a method by which performance data of the combination of rectifier and motor have been calculated by use of a digital computer, together with substantiating data and an example of their use, are included.

A typical bridge rectifier motor controller of the type to be discussed is shown in Fig. 1. The electronic circuits which control the operation of the bridge rectifier provide the functions of linear acceleration rate, armature current limit, and current-resistance drop compensation. Other possible features include the use of tachometer feedback for precise control of speed, and direct control of torque or tension rather than speed.

Bridge Rectifier Circuit

Circuit Description

The bridge rectifier circuit, as used to supply variable voltage and current to the motor armature and constant excitation voltage to the shunt field, is shown in Fig. 2. The alternating supply voltage is generally taken directly from the 230- or 460-volt line. An air-core reactor (or, in some cases, a resistor) is connected in series with the line to limit the amount of short-circuit current which could flow in case of tube backfire or other circuit failure to the maximum allowed by the tube manufacturer's ratings. Tubes *1V* and *2V* are thyratrons whose grids are controlled to vary the armature voltage and current. Gas diode rectifier tubes *3V* and *4V* complete the rectifier bridge circuit.

Circuit Operation

During one half-cycle of the alternating supply the motor armature current is supplied through thyratron *1V* and diode *4V*; during the other half-cycle the armature current is supplied through *2V* and *3V*. There are also two other paths through which armature current flows, as will be shown in a later section. Voltages applied to the two thyratron grids control the instant in their respective half-cycles at which the thyratrons are allowed to conduct, and therefore control the voltage and current applied to the motor armature. These grid control voltages are such that the thyratrons fire at the same point in their respective half-cycles (that is, 180 degrees apart) so that they share the armature current equally.

In addition to their part in supplying power to the motor armature, diode tubes *3V* and *4V* form a half-wave with shunting tube (free-wheeling) rectifier circuit to provide d-c excitation to the motor shunt field. During one half-cycle *4V* conducts and the supply voltage is impressed across the field. During the other half-cycle, the voltage induced in the shunt field self-inductance causes shunt diode *3V* to conduct and thereby maintains the flow of field current. Due to the large value of shunt field inductance, the current thus produced is relatively constant, containing only a small amount of ripple.

Paper 58-40, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 17, 1957; made available for printing November 6, 1957.

R. W. Pfaff is with the Reliance Electric and Engineering Company, Cleveland, Ohio.

The author acknowledges the help and encouragement of Dr. C. G. Veinott and Miss Marilyn Drake, both of Reliance Electric and Engineering Company, in programming the digital computer for the calculation of rectifier characteristics.

Armature Control

When a d-c shunt-wound motor is operating at constant speed with a constant field flux, its equivalent armature circuit consists of a resistance R_a, an inductance L_a, and a constant counter electromotive force (cemf) E_e generated by the rotation of the armature conductors through the field flux. In the following description of circuit operation, the effects on the rectifier of the current drawn through the motor shunt field will be ignored. The motor speed (and correspondingly, voltage E_e) will be considered constant over the period of a half-cycle. The time interval of one half-cycle is so short compared to the response time of the motor that this is a reasonable assumption. Time will be expressed in radian measure with respect to the alternating supply voltage where

$$\theta = 2\pi f t \qquad (1)$$

Fig. 3 shows voltage waveforms observed in the armature circuit. These show the changes in waveform which occur as the load on the motor is increased while maintaining a constant thyratron firing angle. Fig. 3(A) shows the armature voltage waveform when the motor is under light load. A sine-wave supply voltage having a value of E_L volts rms and frequency f is applied to the circuit. At time θ_f, thyratron *1V* is fired by means of its grid control circuit. Supply voltage E_L is then applied to the equivalent circuit shown in Fig. 4(A), consisting of L_{acr} (the inductance of the air core reactors), and tubes *1V* and *4V*, R_a, L_a, and E_e. The voltage drops across *1V* and *4V* during conduction will be assumed to be constant and will be combined into the single term E_{td}. That is, E_{td} is the sum of the voltage drops across one thyratron and one diode.

The voltage appearing at the motor armature terminals while tubes *1V* and *4V* are conducting is equal to the supply voltage less tube drop E_{td} and the relatively small voltage across L_{acr}. During the time the tubes are not conducting, the armature voltage is equal to E_e. The case shown in Fig. 3(A) is of discontinuous conduction. That is, the current flows in pulses separated by intervals of zero current. The differential equation describing the action is

$$2\pi f(L_a + L_{acr})\frac{di}{d\theta} - R_a i - (E_e + E_{td}) = \sqrt{2}E_L \sin\theta \qquad (2)$$

So long as the armature voltage waveform is similar to that of Fig. 3(A), but not Fig. 3(B) or 3(C), operation of the rectifier is quite similar to that of a full-

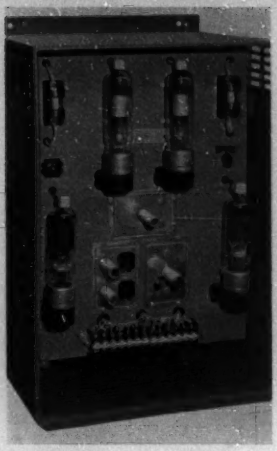


Fig. 1. Typical motor controller using phase-controlled bridge rectifier

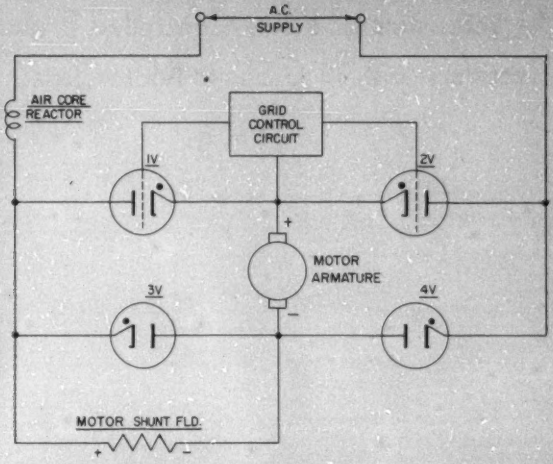


Fig. 2. Phase-controlled bridge rectifier circuit for supplying armature and field of d-c shunt motor

wave rectifier and the formulas shown in reference 1, suitably interpreted, may be used.

In Fig. 3(B) the load on the motor has been increased while the firing angle is held constant. Due to the additional load, E_c has decreased and the tubes carry current for a longer period of time, although the current is still discontinuous.

At the instant labelled θ_d a new mode of rectifier operation is started. At time θ_d the armature voltage is negative, and of a magnitude slightly greater than arc drop E_{td}. When the magnitudes and polarities of the voltages around the series loop consisting of $1V$, the motor armature, and $3V$ are considered it will be seen that the anode of $3V$ has become positive with respect to its cathode, therefore it will conduct current. $4V$ then ceases conducting. Armature current is sustained by the energy which has been stored in the armature inductance, and flows around the loop circuit consisting of $1V$, the armature, and $3V$. The equivalent circuit is shown in Fig. 4(B). This mode of operation continues until the current goes to zero at time θ_e.

While the rectifier is operating in this mode, the voltage at the motor terminals is equal to E_{td}, and the following equation applies:

$$2\pi f L_a \frac{di}{d\theta} - R_a i - (E_c + E_{td}) = 0 \qquad (3)$$

During the interval from θ_e to $(\theta_f + \pi)$, no current flows through the armature and the armature voltage is equal to E_c.

When the rectifier is operating as shown in Fig. 3(B), therefore, it is operating through two distinct modes. In the interval from θ_f to θ_d, current flows to the motor from the supply through $1V$, $4V$, and the air-core reactor (mode I), and equation 2 applies. In the interval from θ_d to θ_e current flows around the loop composed of $1V$, $3V$, and the armature (mode II), and equation 3 applies. The initial condition for equation 3 is that the current immediately after θ_d must be equal to the current just prior to θ_d.

Fig. 3(C) shows the load still further increased. The armature current has not reached zero at the time the following thyratron is fired. The current through the armature does not go to zero at any time during the cycle, this condition being known as continuous conduction. The circuit operates in mode I from θ_f to θ_d, and in mode II from θ_d to $(\theta_f + \pi)$.

When the rectifier is in continuous conduction, the initial condition is imposed on equation 2 that the current just after θ_f be equal to that just before θ_f (which is the same current as exists at $\theta_f + \pi$). This is contrasted to the conditions shown in Figs. 3(A) and 3(B) where the initial condition is that current at time θ_f is zero.

The foregoing description has covered only one half-cycle of operation. Description of the other half-cycle would be identical if $2V$ were substituted for $1V$, and $3V$ for $4V$.

Digital Computer Calculation of Armature Current

Equations 2 and 3, together with the knowledge that during nonconducting intervals the armature voltage is equal to E_c, give a complete mathematical description of rectifier operation. Although solutions to these equations exist, they are unwieldy, do not provide explicit solutions for average or effective (rms) armature currents, and do not seem suitable for practical computational use.

It is desirable to be able to calculate performance data such as average, effective, and peak armature currents for various values of E_c, corresponding to particular values of motor speed, when the parameters R_a, L_a, L_{acr}, E_{td}, and E_L of the motor and bridge rectifier are known.

A straightforward method of calculation which utilizes the high speeds of modern digital computers is used. Equations 2 and 3 are rearranged:

Mode I:

$$\Delta i = \frac{\sqrt{2}E_L \sin\theta - (E_c + E_{td}) - R_a i}{2\pi f (L_a + L_{acr})} \Delta\theta \qquad (4)$$

Mode II:

$$\Delta i = \frac{-(E_c + E_{td}) - R_a i}{2\pi f L_a} \Delta\theta \qquad (5)$$

The integrations implicit in equations 4

and 5 are performed numerically. The time intervals, $\Delta\theta$, of the numerical integration are made sufficiently small that an adequately small error results.

There are three conditions which the computer program must take into account: (1) rectifier operating in mode I where equation 2 applies, Fig. 3(A); (2) rectifier operating through both modes I and II where both equation 2 and equation 3 must be used, Fig. 3(B); and (3) rectifier operating in modes I and II and also in continuous conduction, Fig. 3(C), where the computer must also determine the current at time θ_f, when *1V* is fired.

To make the computation, constants R_a and L_a of the motor and E_L, f, L_{ac}, and E_{td} of the rectifier must be known. Values of E_c and θ_f are picked, and the step-by-step calculation is carried out. The incremental change in armature current which occurs during a time interval is calculated using equations 4 or 5. The total armature current at the end of the nth time interval is

$$i_n = i_{n-1} + \Delta i_n \tag{6}$$

where i_{n-1} is the current at the end of the $(n-1)$th interval, and Δi_n is the change in current during the nth interval.

These computations are carried forward until the current i_n has fallen to zero, or until $(\theta_f+\pi)$ is reached. The average armature current is then

$$I_{avg} = \frac{1}{\pi}\sum i_n \Delta\theta \tag{7}$$

As simple, but very valuable, by-products of the computations, the effective armature current

$$I_{rms} = \sqrt{\frac{1}{\pi}\sum i_n^2 \Delta\theta} \tag{8}$$

the peak armature current, and the conduction angle (time over which the thyratron remains conducting) may be obtained.

The computation is carried out as follows:

1. Starting at time θ_f with an assumed armature current of zero, calculate the first increment of current:

$$\Delta i_1 = \frac{\sqrt{2}\,E_L \sin\left(\theta_f + \frac{\Delta\theta}{2}\right) - (E_c + E_{td})}{2\pi f(L_a + L_{ac})}\,\Delta\theta$$

The current at the end of the first increment of time is $i_1 = \Delta i_1$.

2. Calculate the second increment of current as

$$\Delta i_2 = \frac{\sqrt{2}\,E_L \sin\left(\theta_f + \frac{3\Delta\theta}{2}\right) - (E_c + E_{td}) - R_a i_1}{2\pi f(L_a + L_{ac})}\,\Delta\theta$$

The current at the end of the second increment of time is $i_2 = i_1 + \Delta i_2$.

3. Continue the foregoing until θ_d is reached.

4. For the period from θ_d to θ_e, or from θ_d to $(\theta_f+\pi)$ if the rectifier is found to be in continuous conduction, equations 5 and 6 are modified and used.

5. If the rectifier is not in continuous conduction, i_n will be found to be zero at the end of one of the calculations in steps 3 or 4. Calculation of motor currents will then be made as indicated in step 7.

6. If the rectifier is in continuous conduction, current i_n will be greater than zero when $(\theta_f+\pi)$ is reached. A new value of current at time θ_f will be assumed (greater than zero) and the computations described will be repeated. This will follow through as an iterative process until an initial current at θ_f is found which is the same as the current thereafter computed at $(\theta_f+\pi)$, thus satisfying the initial conditions.

7. At the end of the foregoing computations, the average and effective values of armature current are computed using equations 7 and 8. The peak value of armature current is stored in the computer during the course of the computation, and the conduction angle is computed as

$$\theta_k = \theta_e - \theta_f$$

All of the foregoing values are printed out in a convenient tabular form. The medium-speed digital computer which is used makes the calculations indicated and prints out a set of computed data

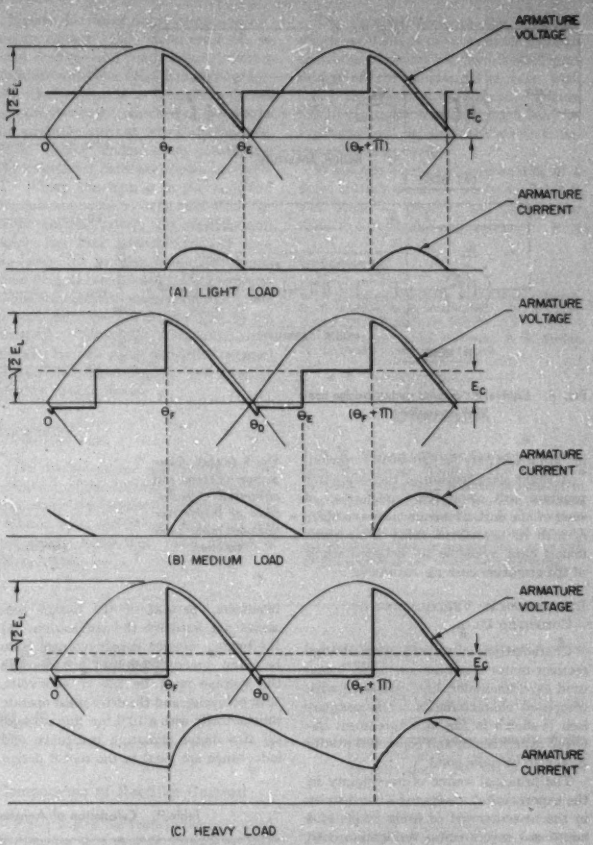


Fig. 3. Typical armature voltage and current waveforms

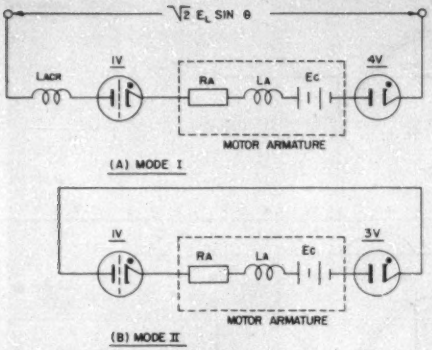


Fig. 4. Equivalent circuits during bridge rectifier operation

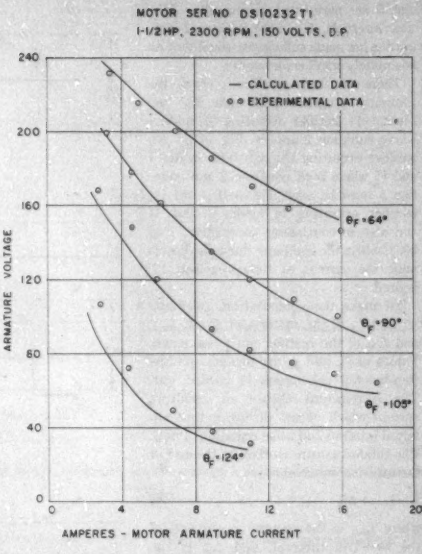


Fig. 5 (right). Comparison of actual and computed characteristics of bridge rectifier–d-c motor combination

in approximately a 20-second period.

A slight modification of the computer program will allow the printing-out of each of the instantaneous current values i_a with its associated value of θ_a, which makes data available for detailed study of the armature current waveform.

Experimental Verification of Computed Data

Characteristic curves of a sample bridge rectifier-motor combination were measured experimentally, and compared with computed characteristics. The comparison is shown in Fig. 5. Agreement between experimental and computed results is seen to be quite good.

The principal source of uncertainty in the experimental results is believed to be in the measurement of firing angle, and additional experimental verification is to be made in the near future.

Computer Study of Bridge Rectifier Voltage Ratings

Digital computer calculation of rectifier drive characteristics has been used in determining the commercial voltage ratings of the bridge rectifier–motor combinations. Several of the design constants are fixed by the application: (1) the bridge rectifier supply voltage is to be taken directly from the line, therefore the voltage must be 230 or 460 volts, 50 or 60 cycles, and the drive must operate satisfactorily with a 10% low line voltage; (2) the motor armature resistance and inductance are fixed in the motor design and may not easily be altered by the rectifier designer (adding resistance or inductance in series with the armature lowers the maximum rectifier output voltage under load); (3) rectifier tube characteristics such as arc drop and starting voltage are usually beyond the rectifier designer's control.

Armature voltage ratings had been

Table I. Minimum Firing Angles

Actual Line Voltage	Armature Voltage Rating	Minimum Firing Angle, Degrees (Minimum θ_f)
414	320	43.6
460	320	38.5
207	160	59.3
230	160	50.8

Table II. Calculation of Armature Current Under Various Conditions

Motor Ratings: Horsepower	Rpm	Volts	Amperes	R_a (Ohms)	L_a (Millihenrys)	L_{acr} (Millihenrys)	Frequency (Cps)	Rectifier Output Current, Amperes: 414-Volt Line	460-Volt Line
3	2,500	320	7.9	1.53	24.5	1.8	60	14.5	21.1
3	1,750	320	7.9	1.9	34.8	1.8	60	10.8	19.0
2	2,500	320	5.4	2.5	30	1.8	60	11.4	16.5
2	1,750	320	5.4	2.67	42.3	1.8	60	8.7	13.6
2	1,150	320	5.4	4.39	77	1.8	60	5.3	10.6
1½	2,500	320	4.2	3.5	44.8	1.8	60	7.95	11.5
1½	1,750	320	4.2	4.34	52	1.8	60	6.96	10.0
1½	1,150	320	4.2	6.12	98	1.8	60	4.19	7.95
1	2,500	320	2.8	6.3	54.5	1.8	60	6.09	8.71
1	1,750	320	2.8	6.05	77.6	1.8	60	4.73	6.85
1	1,150	320	2.8	10	120	1.8	60	3.29	5.16
1	850	320	2.8	8.05	143	1.8	60	2.87	5.72
								207-Volt Line	230-Volt Line
2	2,500	160	10.8	0.627	7.5	0.9	60	15.4	24.7
2	1,750	160	10.8	0.67	10.6	0.9	60	11.8	19.0
2	1,150	160	10.8	1.1	19.2	0.9	60	7.30	11.7
1½	2,500	160	8.4	0.875	11.2	0.9	60	10.8	17.4
1½	1,750	160	8.4	1.08	13	0.9	60	9.56	15.5
1½	1,150	160	8.4	1.53	24.5	0.9	60	5.78	9.24
1	2,500	160	5.6	1.57	13.6	0.9	60	8.49	13.4
1	1,750	160	5.6	1.51	19.4	0.9	60	6.54	10.5
1	1,150	160	5.6	2.5	30	0.9	60	4.58	7.25
1	850	160	5.6	2.01	35.8	0.9	60	3.96	6.38

tentatively set at 320 volts for operation from a nominal 460-volt line, and 160 volts with a nominal 230-volt line. The computer was used to determine whether the bridge rectifier could supply such rated voltages at rated load over a line of motors, and the 10% low line condition.

Study of tube manufacturers' ratings showed that a reasonable value of thyratron starting voltage is 100 volts (*NL-740, NL-760, C6J/K, GL-6807*), and that a suitable value for E_{td} is 24 volts. As noted before, E_{td} is the sum of the arc drops of one thyratron plus one diode.

The minimum firing angle, θ_f, is set by the values of the supply voltage, E_L; thyratron starting voltage, E_{td} and motor cemf, E_c, as indicated in the following:

$$\text{Minimum } \theta_f = \sin^{-1} \frac{E_c + E_{td}}{\sqrt{2}\, E_L}$$

For the sizes of motors to be used with these drives (1/2-3 horsepower), the armature cemf, E_c, at rated load and voltage is about 95% of the armature voltage rating. The minimum firing angles are therefore as shown in Table I. Values of air-core reactor inductance (L_{acr}) as shown in Table II were used.

Figures for firing angle (minimum θ_f), supply voltage (E_L), motor cemf (E_c at rated voltage and load current), motor armature resistance (R_a), inductance (L_a), and other constants as indicated were set into the computer, which then calculated the armature current which the rectifier would supply under those conditions. The calculated data are shown in Table II. Where the figures in the last two columns are equal to or greater than the motor current rating, the rectifier will supply the load satisfactorily. If they are lower, the rectifier would not supply rated load at rated voltage to the motor.

The calculations summarized in Table II were performed to supply data to the National Electrical Manufacturer's Association, which is presently engaged in standardizing ratings for these bridge rectifier motor drives.

Summary

The mathematical description of the operation of the rectifier-motor combination, and its reduction to automatic computation by a digital computer, will permit greater scope and closer design of these drives.

The detailed data on the armature which are made readily available through use of the computer will permit intelligent design of the motors. The ease with which the effects of changes in armature circuit resistance and inductance may be evaluated will be an aid when trial designs are being made.

When the over-all characteristics of a drive system are made easily available it will often be possible to "tailor" the drive to a particular application.

References

1. Theory of Rectifiers—D-C Motor Drives, R. H. Vedder, K. P. Puchlowski. *AIEE Transactions*, vol. 62, 1943, pp. 863-70.

2. Thyratron Motor Control, E. E. Moyer, H. L. Palmer. *Ibid.*, pp. 706-12.

Discussion

J. T. R. Wilson (Aerojet General Corporation, Sacramento, Calif.): The subject is very well presented. A point of interest to me which was not fully covered was the grid control circuit used. It is understood that various types of grid control circuits might be used but an example of one or two or the preferred one would be interesting.

Power Factor of Rectifiers

A. SCHMIDT, JR.
FELLOW AIEE

THE POWER FACTOR of an electric apparatus is a means of relating kilowatts to apparent power or kilovolt-amperes. With sinusoidal currents, its value permits determination of reactive power, which is a factor in defining transmission losses, regulation, and generator excitation. The effect of reactive power is often reflected in added user charges for low power factor. These charges usually justify the cost of improving power factor.

If the current is nonsinusoidal, the numerical value of power factor loses some of its significance. Because of this, the section on power factor in the forthcoming revision of the American Standards for mercury-arc power converters[1] has been modified to give less weight to harmonic components of rectifier input current. This paper is a discussion of rectifier power factor and methods for determining its value. The principles involved may be applied to other devices with nonsinusoidal current.

Components of Rectifier Current

Fig. 1 shows the transformer a-c winding current in a double 3-phase rectifier both for the idealized case and for the condition of commutating reactance and phase control. In both cases, high d-c reactance is assumed. The current has an rms value of[2]

$$I_p = \sqrt{\frac{2}{3}}\, I \sqrt{1 - c f(u, \alpha)} \qquad (1)$$

where

I = amplitude of transformer a-c winding current

Paper 58-220, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing November 29, 1957.

A. Schmidt, Jr., is with the General Electric Company, Schenectady, N. Y.

The author acknowledges the suggestions made in a discussion by Kevin F. Frederick, Westinghouse Electric Corporation, which have been incorporated in this paper.

I_p = rms value of transformer a-c winding current

$$f(u, \alpha) = \frac{1}{2\pi}\left(\frac{\sin u\,[2 + \cos(u + 2\alpha)] - u\,[1 + 2\cos\alpha\cos(u + \alpha)]}{[\cos\alpha - \cos(u + \alpha)]^2}\right) \qquad (2)$$

c = a constant whose value depends on the type and part of the circuit under consideration. For the 3-phase circuit, $c = 3$

u = commutating angle

α = angle of phase control

The quantity $\sqrt{1 - cf(u, \alpha)}$ is a measure of the effect of commutation on the waveform and is termed the overlap factor.

This current can be broken down into a fundamental power component, a fundamental reactive component, and a series of harmonic components. Their values are given by:

$$a_1 = \frac{\sqrt{3}\, I}{\sqrt{2}\, \pi}\,[\cos\alpha + \cos(u + \alpha)] \qquad (3)$$

$$b_1 = \frac{\sqrt{3}\, I}{\sqrt{2}\, \pi}\,\frac{\sin(2u + 2\alpha) - \sin 2\alpha - 2u}{2[\cos\alpha - \cos(u + \alpha)]} \qquad (4)$$

$$I_H^2 = I_p^2 - a_1^2 - b_1^2 \qquad (5)$$

where

a_1 = effective value of fundamental power component of I_p

b_1 = effective value of fundamental reactive component of I_p

I_H = effective harmonic component of I_p

Fig. 2 shows the effective values of I_p and its components as a function of u and α.

Power Factor

The power factor is the ratio of real to apparent power. It may be expressed

$$\text{pf} = \frac{a_1}{I_p} \tag{6}$$

This value is not a true measure of system conditions due to the rectifier load.

1. If a-c system regulation is calculated without taking into account the effect of harmonic current on power factor, the value will be too high because the calculated wattless current will be too high.

2. Power-factor corrective means are not effective for harmonic components and these components should not be included in correction calculations.

3. Local heating in generator fields due to harmonic components in the armature current may be out of proportion to the magnitude of the harmonic current components. Generator excitation is determined by current displacement, and the heating due to harmonics is determined separately, taking into account the magnitude and order of the harmonic currents.

4. Conductor losses are higher than indicated by effective current values because of the skin effect at harmonic frequencies. This is recognized by transformer designers who commonly design rectifier transformers on the basis of the idealized rectangular current wave (A) instead of the practical wave (B) shown in Fig. 1. The design for overcurrents compensates for the added high-frequency losses.

Displacement Power Factor

Displacement power factor is the value of power factor arrived at by considering only the fundamental components of current. The value is given by

$$\cos\phi = \frac{a_1}{\sqrt{a_1^2 + b_1^2}} = \cos\left(\tan^{-1}\frac{b_1}{a_1}\right) \tag{7}$$

This value will be modified by the exciting current of the rectifier transformer and of phase-shifting transformers, regulating transformers, and induction regulators, if used. The modified value is

$$\cos\phi_1 = \frac{a_1}{\sqrt{a_1^2 + (b_1 + I_e)^2}} = \cos\left(\tan^{-1}\frac{b_1 + I_e}{a_1}\right) \tag{8}$$

where

$\cos\phi$ = displacement power factor of rectifier, neglecting transformer exciting current

$\cos\phi_1$ = displacement power factor of rectifier, including effect of exciting current

I_e = the total exciting current

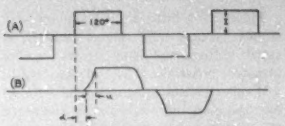


Fig. 1. Transformer a-c winding current in double 3-phase rectifier

A—Idealized waveform with no reactance or phase control

B—Effect of commutating reactance and phase control

The simplifying assumption is made that all exciting currents are sinusoidal and purely reactive. This assumption leads to calculated values which are very slightly low.

Measurement of Power Factor

If meters are used to measure volts, amperes, and watts in the supply to the rectifier, power factor is determined by the following:

$$\text{pf} = \frac{W}{\sqrt{3}E_L(I_L)} \tag{9}$$

This is equivalent to the value expressed in equation 6.

Metering arrangements involving the use of power-factor meters or combinations of wattmeters, varmeters, and volt-ampere meters will yield values of displacement power factor, rather than power factor, because such meters generally integrate only the product of volts and amperes of the same frequency. Thus, any practical method of indicating power factor for billing charges must be based on displacement power factor. It follows that the application of power-factor corrective means should have this basis.

Calculation of Power Factor

The direct measurement of power factor or displacement power factor is subject to error because of distortion of the line voltage by the system reactance and because of instrument inaccuracies at harmonic frequencies. For these reasons and also because of the difficulty in making field tests under controlled conditions and the expense of factory tests, standardizing bodies have generally adopted calculation as the method of determining power factor and displacement power factor.

The foregoing equations may be used to determine power factor and displacement power factor. Information presented in terms of u and α is not convenient for general use. By using the following generalized equations for a_1 and b_1 rather than equations 3 and 4,

$$a_1 = \frac{\sqrt{2}}{\pi} I \sin\frac{\pi}{p}\left[\cos\alpha + \cos(u+\alpha)\right] \tag{10}$$

$$b_1 = \frac{\sqrt{2}}{\pi} I \sin\frac{\pi}{p} \times \left(\frac{\sin(2u+2\alpha) + \sin(2\alpha-2u)}{2[\cos\alpha + \cos(u+\alpha)]}\right) \tag{11}$$

where p is the number of phases in a commutating group, and noting that

$$\frac{E_x}{E_{do}} = {}^1/_2[\cos\alpha - \cos(u+\alpha)] \tag{12}$$

where E_x is the d-c drop in the rectifier due to commutating reactance and E_{do} is the theoretical no-load voltage of the rectifier, displacement power factor may be presented in terms of per-unit d-c drop due to reactance, and in this form this value is independent of the number of rectifier phases.

Fig. 3 shows calculated displacement factor presented in this manner. Conversion factors for converting commutating reactance to Ex/E_{do} at rated load are given in Table I for the most commonly used circuits. At other loads Ex/E_{do} is proportional to load, except at high values of reactance factor. If there is reactance between the rectifier transformer and the point at which power factor is to be measured, this reactance should be added to the transformer reactance to determine Ex/E_{do}, first multiplying the line reactance by a transformation constant whose value depends on the transformer connection.[1]

A different procedure should be followed if several phase-displaced rectifier transformers are connected to a common bus and there is reactance between this bus and the point at which power factor is to be measured. In this case, there is usually sufficient phase multiplication so that at least two, and usually three, rectifiers have overlapping commutation, and the transformer terminal voltage can be considered sinusoidal because of the nearly sinusoidal drop in the common line reactance. The displacement power factor of one of the rectifiers may thus be determined on the basis of sinusoidal applied voltage. The power factor at the measuring point is the resultant of the power factor of the individual rectifiers, the

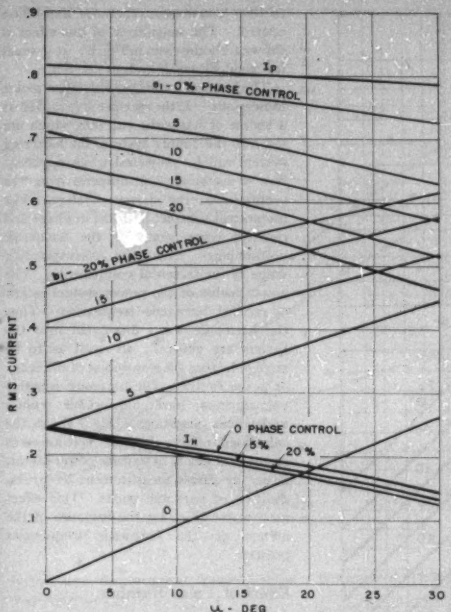


Fig. 2. Components of transformer a-c winding in Fig. 1. Per-unit base is amplitude of winding current

I_p = rms value of current
a_1 = fundamental effective in-phase component of I_p
b_1 = fundamental effective reactive component of I_p
I_H = effective harmonic component of I_p

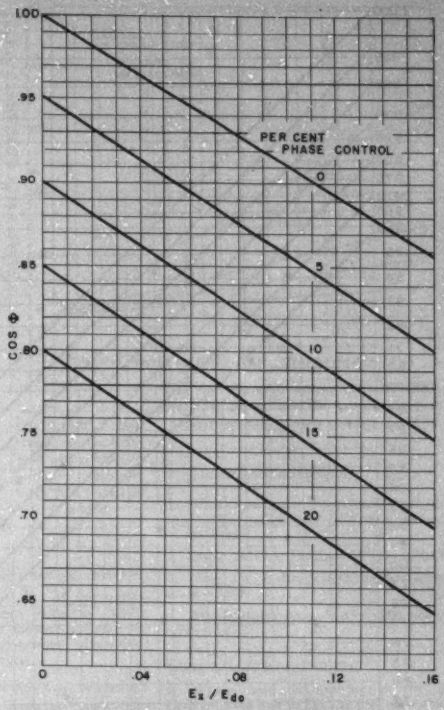


Fig. 3. Displacement power factor cos ϕ of rectifier. E_x/E_{do} is the per-unit drop in direct voltage due to commutating reactance

common line reactance, and the total rectifier load.

Fig. 4 permits the conversion from cos ϕ to cos ϕ_1 when the exciting currents of the rectifier transformer and regulating and phase-shift transformers (if present) are known.

Approximate Method for Calculation of Displacement Power Factor

Fig. 3 shows the displacement power factor for the range of conditions usually encountered. When large amounts of phase control are present, the following approximate expression may be used:

$$\cos \phi = 1 - Ex/E_{do} - \frac{\%\ \text{phase control}}{100} \qquad (13)$$

Thus, a 6-phase double-Y rectifier with 10% commutating reactance (Ex/E_{do} = 0.05) and 30% phase control has an approximate displacement power factor of

$$\cos \phi = 1 - 0.05 - 0.30 = 0.65$$

at rated load.

The error is small for phase control in excess of 20%.

Effect of System Reactance

The foregoing analysis is based on pure sinusoidal applied voltage. Since there is always reactance in the supply system, its effect on power factor should be noted.

Reactance in the supply system distorts the voltage at the rectifier terminals, as illustrated in Fig. 5. One effect of the voltage distortion at the rectifier terminals is a change in displacement power factor measured at that point. This may seem strange since the power factor of conventional devices taking sinusoidal current is not altered by the insertion of reactance between the measuring point and the supply system. This situation may be rationalized by considering the effect on power factor when inductance is inserted between an infinite bus and a rectifier. The reduction in displacement power factor at the bus varies directly with the inductance. This is illustrated in Fig. 3, bearing in mind that at a given

Table I. Rectifier Circuit Factor

ASA Circuit No.	Name of Circuit	K_x Where $Ex/E_{do} = K_x \times$ Per-Unit Commutating Reactance	
		Transformer	Line
45	6-phase double-Y	0.5	0.5
25	6-phase double-way	0.5	0.5
43	6-phase triple diametric	0.75	0.5
2	Diametric	0.707	0.707
11	6-phase fork (star)	0.5	0.5

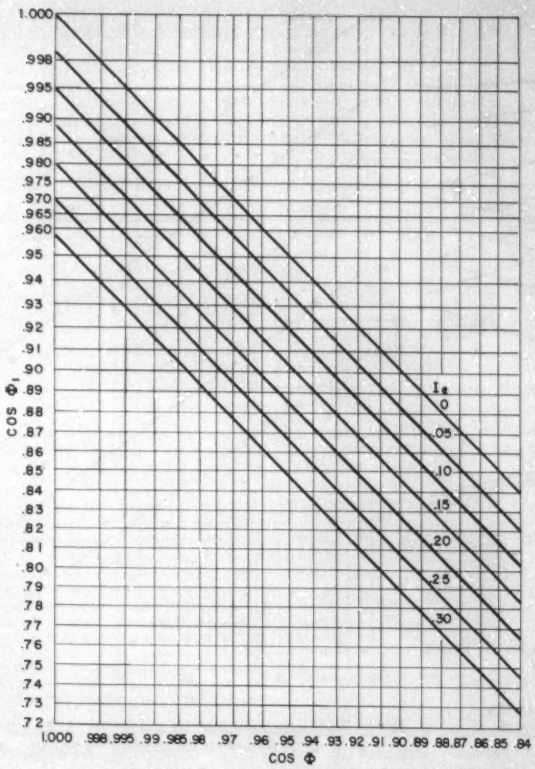


Fig. 4. Displacement power factor cos ϕ_1, showing correction for transformer exciting current I_e. Per-unit base of exciting current is power component of input current at load under consideration

current Ex/E_{do} will vary directly with the total inductance. If the rectifier is replaced with a resistor, the reduction in power factor of the combination will vary not directly but with the square of the inductance for moderate values of reactance, say up to 20 or 30%, which is the upper limit of reactance usually encountered in rectifier circuits.

Since a change in power factor with line reactance is expected, it is desirable to determine the magnitude of this change. The displacement power factor at the rectifier terminals with reactance in the supply line may be determined by one of two methods:

1. The displacement power factor may be determined at the power source, treating the system reactance as a part of the commutating reactance. From the displacement factor, the power and reactive components of current are derived and the displacement angle ϕ at the rectifier terminals is calculated from the drop in the line reactance due to these components. This is a conventional procedure.

2. The displacement angle ϕ at the power source may be determined as in method 1. Then the voltage wave at the rectifier terminals (Fig. 5) may be analyzed to determine its displacement from the voltage at the power source. This displacement is subtracted from ϕ at the source to determine the displacement angle at the rectifier terminals.

Displacement power factor in Fig. 3 is the value at the rectifier terminals based on zero line reactance. The effect of line reactance is to decrease this value by an amount depending on line reactance, rectifier transformer reactance, and phase control. The magnitude of this effect is shown for some cases in Fig. 6. It is small for usual values of system reactance.

The harmonic voltages also affect meter indications. If the rectifier is regarded as a source of harmonic currents which are fed into the supply system, the metering system will be subjected to an apparent flow of harmonic volt-amperes from the rectifier to the power system. The meters will subtract both the in-phase and reactive components of the harmonic volt-amperes from the corresponding value of fundamental components. The power factor of the power system is low at rectifier harmonic frequencies. Thus the harmonic watts dissipated into the system are generally so small as to be negligible from the standpoint of efficiency or power factor. The harmonic reactive volt-amperes have appreciable values under some conditions. Fig. 7 shows the calculated effect of these volt-amperes on meters used to determine power factor, either by direct measurement or by indication of vars and watts. This effect may be modified by the response of the meters at the harmonic frequencies present.

Effect of Pulse Number

The pulse number or number of rectifier phases is generally six or a multiple of six in equipments with ratings of more than a few kilowatts. While the large d-c inductance postulated in rectifier circuit analysis is seldom present, the current harmonics in practical circuits do not differ greatly from the theoretical values in 6-phase rectifiers. This difference decreases with increasing pulse numbers.

The only important application of single-phase, or diametric, rectifiers with a pulse number of two, is in the traction field. In this application, the d-c circuit inductance is limited and the departure from ideal waveforms is considerable. A large third harmonic is present so that the harmonic content of the line current is considerably greater than in polyphase systems. The spread between power

Fig. 5. A-c line voltage waveform at line terminals of 6-phase double-way rectifier transformer, showing distortion due to line reactance; 7½% transformer reactance, 2½% line reactance, 14% phase control

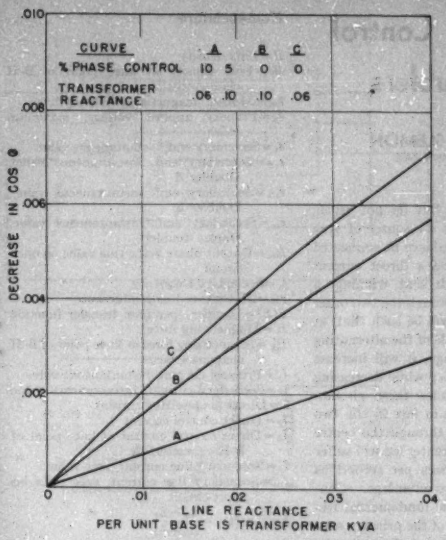


Fig. 6 (left). Decrease in displacement power factor at rectifier terminals when line reactance is added

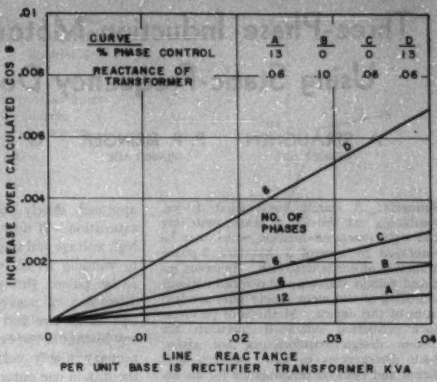


Fig. 7. Apparent increase in metered displacement power factor due to action between rectifier harmonic currents and line reactance

factor and displacement power factor may approach 10%.

Rectifier Input Volt-Amperes

Three different values of volt-amperes are associated with the rectifier input, having numerical values which approach each other as the number of rectifier phases increases. They are

1. The fundamental component of volt-amperes taken from the power line. Its numerical value is input watts divided by displacement power factor. The value can be measured with a volt-ampere meter. It can also be calculated from measurements with a wattmeter and a varmeter, or a wattmeter and a power-factor meter.

2. The total volt-amperes taken from the line. Its numerical value is input watts divided by power factor. The value can be measured with a volt meter and ammeter.

3. The volt-ampere rating of the transformer. Its numerical value is input watts divided by the theoretical value of power factor with zero overlap. Its value is the highest of the three.

The relative values of volt-amperes, using the first value as a base, are shown in Table II.

Example

As an example, consider the case of a 1,000-kw 6-phase double-Y rectifier, having measured losses of 80 kw with the transformer having 7% commutating reactance and an exciting current of 3% on the transformer kilovolt-ampere base. The transformer is compensated for 5% drop in line voltage, so that the rectifier will deliver rated voltage and current at 95% rated alternating volts.

1. If rated alternating voltage is supplied, the rectifier will require 5% phase control at rated load.

2. From Table I, $Ex/E_{do} = 0.07/2 = 0.035$.

3. From Fig. 3, displacement power factor $\cos \phi = 0.918$.

4. Fundamental input kva $= (1{,}000 + 80)/0.918 = 1{,}176$.

5. From Table II, the transformer kva rating is $1{,}176 \times 1.047 = 1{,}231$.

6. If the base for exciting current is changed from kva to kw, the per-unit value of I_e becomes $0.03 \times 1{,}231/1{,}080 = 0.034$.

7. From Fig. 4, $\cos \phi_1 = 0.906$.

8. The kva taken from the line, including harmonics, is indicated by Table II to be $1{,}176 \times 1.025 = 1{,}205$.

Table II. Relative Values of $\frac{\text{Volt-Amperes}}{\text{Fundamental Component of Volt-Amperes}}$

Method	Number of Rectifier Phases 2	6	12
1. Volt-amperes measured with Volt-ampere meter; Wattmeter and varmeter; Wattmeter and power-factor meter	1.0	1.0	1.0
2. Volt-amperes measured with volt meter and ammeter (typical reactance values)	1.09*	1.025†	1.01
3. Volt-ampere rating of transformer a-c winding	1.111*	1.047	1.01†

* May vary appreciably with low d-c inductance.
† Will vary ±0.01 with typical values of reactance and phase control.

Summary

Power factor and the displacement component of power factor of a rectifier are derived from the components of the current to the rectifier. Displacement power factor is presented on the basis of circuit reactance, with correction for exciting currents. Line reactance affects both the indications of instruments and the value of power factor. The result of these variations is small for usual line reactance values.

References

1. American Standard for Pool Cathode Mercury-Arc Power Converters. *ASA C-34.1-1949*, American Standards Association, New York, N. Y., 1949.

2. Analysis of Rectifier Circuits, E. F. Christensen, C. H. Willis, C. C. Herskind. *AIEE Transactions*, vol. 63, 1944, pp. 1048–58.

Three-Phase Induction-Motor Control Using Static-Frequency Doublers

A. STRAUGHEN MEMBER AIEE P. P. BIRINGER MEMBER AIEE G. R. SLEMON MEMBER AIEE

Synopsis: A simply constructed device, combining the benefits of static switching and static power modulation, which may be employed to control a standard 3-phase induction motor is described. Inherent reduced voltage starting and reversal without switching of the main supply lines are features of the device. Methods of performance prediction sufficiently accurate for system design requirements are given. Such divergencies as exist between predicted and realized performance for an experimental system consisting of device and standard motor are explained.

THE METHOD of driving a 2-phase induction motor by means of static frequency doublers has been described in reference 1. The extension of that method to a 3-phase induction motor drive of integral horsepower rating is described in the following.

The operation of the static frequency doubler had been described in reference 2 and a method of design has been given in reference 3. Here it will suffice if a very brief sketch of the behavior of a single static frequency doubler is given, followed by a fuller explanation of the duplex connection originally discussed in reference 1.

Single Static Frequency Doubler

Fig. 1 shows the winding arrangement and corresponding diagram of connections for the type of static frequency doubler whose use is described in the following. An alternating voltage source is connected to terminals *P*1 and *P*3 and produces an alternating flux in the magnetic circuit formed by the two outer legs of the core and the yokes. With the windings on the center leg of the core open circuited, no flux passes through the centre leg, and hence no electromotive force (emf) is induced in the windings on that leg. The doubler is so designed that under these conditions the two outer legs approach closely to, but do not reach, saturation. If now a d-c source of very high voltage and impedance be connected to winding *C*1-*C*2 and a direct current I_c be passed through that winding, a direct flux will pass through the two outer legs of the core and will be such that, at any instant in the cycle of the alternating primary supply voltage, it will increase the flux in one outer leg while decreasing it in the other. Under these circumstances the difference in flux in the two outer legs must pass through the centre leg. This flux in the center leg will suffer twice as many reversals per second as does the flux in the two outer legs. Consequently, emf's whose fundamental frequency are twice that of the primary supply are induced in the windings on the center leg. The high impedance of the d-c source will prevent the flow of any appreciable alternating current in winding *C*1-*C*2. However, if a load be connected to the terminals of winding *S*1-*S*2 then alternating current will flow in that load.

It has been shown[2] that, if the magnetization characteristic of the core material be idealized to the straight line form shown in Fig. 2, then the emf induced in the secondary winding *S*1-*S*2 when direct current flows in the control winding *C*1-*C*2 has the waveform shown in Fig. 3(A). The magnitude of the angle θ_s shown in that figure depends upon the magnitude of the direct control current. When I_c is zero, θ_s is 90 degrees, and no emf is induced. As I_c is increased to the value I_k indicated in Fig. 2 as just producing saturation of the core, $\theta_s = 0$ and the secondary induced emf is a maximum. As I_c is further increased, θ_s also increases and the secondary emf once again falls. For some value of I_c much greater than I_k, a value which depends on the characteristic of the core material and the construction of the doubler, θ_s again reaches 90 degrees and the secondary emf vanishes. In the practical doubler, the effects of leakage inductance and of the rounded knee of the magnetization characteristic combine to produce a secondary emf whose waveform is much smoother and more rounded than the theoretical waveform shown in Fig. 3(A).

Paper 58-68, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 17, 1957; made available for printing December 10, 1957.

A. Straughen is with Canadian Controllers Ltd., Scarboro, Ont., Canada; P. P. Biringer and G. R. Slemon are with the University of Toronto, Toronto, Ont., Canada.

Nomenclature

B = Flux density
B_k = Flux density at knee point of B-H characteristic
b_m = Motor magnetizing susceptance
e_p = Primary applied voltage, maximum value
e_s = Secondary emf, instantaneous value
e_{sA} = Secondary emf, instantaneous value, Doubler *A*
e_{sB} = Secondary emf, instantaneous value, Doubler *B*
e_{sc} = Secondary emf, instantaneous value, Duplex doubiers
E_{oc} = Output phase volts, rms value on open circuit
f_s = Secondary frequency
g_{hs} = Motor core loss conductance
$G(S)$ = Doubler open-loop transfer function
H = Magnetizing force
H_k = Magnetizing force at knee point of B-H characteristic
i_p = Primary current, instantaneous value
i_s = Secondary current, instantaneous value
I = Direct magnetizing current
I_c = Direct control current
I_k = Direct control current at knee point of B-H characteristic
I_s = Secondary line current, rms value
I_{sc} = Secondary line current, rms value on short circuit
k = constant $= \frac{L_u - L_s}{L_u + L_s}$
L_u = Primary inductance, core unsaturated
L_s = Primary inductance, core saturated
PF = Power factor
PPF = Primary power factor
P_s = Secondary power output
p = Number of stator poles
r_1 = Stator resistance per phase
r_2' = Rotor resistance per phase referred to stator terminals
s = Fractional slip
S = Laplace operator
T = Torque
T_c = Time constant
V_s = Secondary phase volts, rms value
X_{int} = Secondary internal reactance per phase
X_1 = Stator leakage reactance per phase
X_2' = Rotor leakage reactance per phase referred to stator terminals
η = Efficiency
θ = Supply voltage phase angle
θ_s = Switching angle
ϕ = Total flux in a core
ϕ_k = Total flux in a core at knee point of B-H characteristic

Duplex Doubler Connection

The secondary emf of a single frequency doubler shown in Fig. 3(A) may be represented over one half-cycle of the supply frequency by the equations[2]

$$0 < \theta < \theta_s \qquad e_s = 0 \tag{1}$$

$$\theta_s < \theta < \pi - \theta_s \qquad e_s = -\hat{e}_p k(\cos\theta) \tag{2}$$

$$\pi - \theta_s < \theta < \pi \qquad e_s = 0 \tag{3}$$

where

$$\theta_s = \mathrm{Sin}^{-1}\frac{2\omega L_u}{\hat{e}_p}(I_k - I_c) \text{ for } I_c < I_k \tag{4}$$

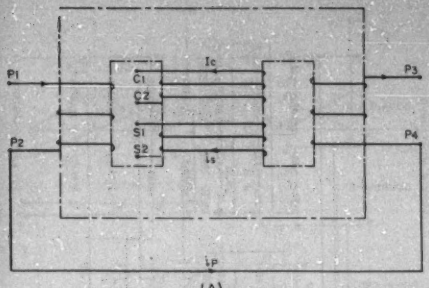


Fig. 1. The static frequency doubler

A — Arrangement of windings
B—Diagram of connections

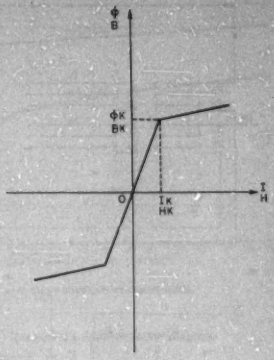


Fig. 2 (right). Idealization of the magnetization characteristic

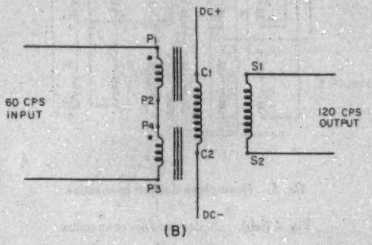
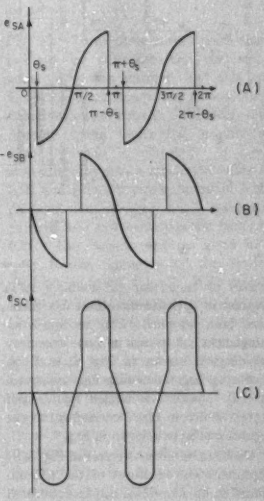


Fig. 3 (right). Emf wave form of duplex - connected doublers

A—Doubler A
B—Doubler B
C—Resultant emf



$$\theta_s = \mathrm{Sin}^{-1}\frac{2\omega L_s}{\hat{e}_p}(I_c - I_k) \quad \text{for } I_c > I_k \qquad (5)$$

The normal condition of operation of the doubler is when $I_c > I_k$.

The duplex doubler connection consists of two doublers whose primaries are supplied from voltage sources 90 degrees out of phase with one another, as a result of which the second harmonic outputs of the two doublers are 180 degrees out of phase as in Fig. 3(A) and 3(B). If the two doubler outputs be now connected in opposition, all 4th, 8th ...4nth harmonics are cancelled in the combined output, with the result that a secondary emf wave form with alternating symmetry is obtained. The theoretical waveform obtained by piecewise linear analysis of the system operation is shown in Fig. 3(C). It is represented over one quarter cycle of the supply frequency by

$$0 < \theta < \theta_s \qquad e_{sc} = -\hat{e}_p k(\sin\theta) \qquad (6)$$

$$\theta_s < \theta < \pi - \theta_s \qquad e_{sc} = -\hat{e}_p k(\sin\theta + \cos\theta) \qquad (7)$$

$$\pi - \theta_s < \theta < \pi \qquad e_{sc} = -\hat{e}_p k(\cos\theta) \qquad (8)$$

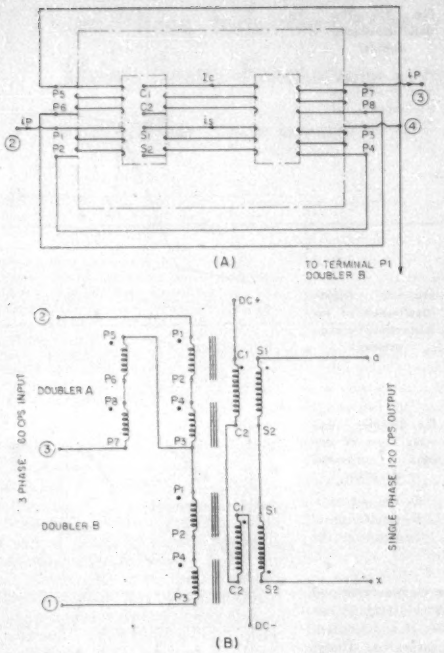
In the practical case, this waveform approximates closely to a sinusoid of second harmonic frequency.

In order to obtain the required primary supply voltages 90 degrees out of phase, the modified Scott connection shown in Fig. 4 was employed. The use of four windings to establish a midpotential point between terminals 2 and 3 is necessitated by the fact that, for part of the cycle, one primary core of doubler *A* is saturated, while the other is unsaturated. Under these conditions almost the entire primary applied voltage appears across the terminals of the windings on the unsaturated core.

3-Phase Doubler Connection

The obvious method of obtaining a controllable 3-phase output by means of doublers would be that of using three duplex pairs and Y-connecting the second harmonic outputs. Such an arrangement does not, however, permit of phase sequence reversal without switching since reversal of the secondary emf of any or all of the duplex pairs does not produce a balanced system of opposite phase sequence. A less direct method, requiring two groups of single-phase outputs in quadrature, is therefore employed. Each group of single-phase outputs is obtained from one pair of duplex-connected doublers, and since these groups of outputs must be in quadrature, it follows that the inputs to the two pairs of doublers producing them must be displaced 45 degrees from one another.

An alternative to phase displacement of the voltages applied to the primary windings of the two pairs of doublers is that of phase displacement of their flux waves by means of additional primary windings. Thus a connection such as is shown in Fig. 4 could be employed to give one group of secondary outputs, and a similar connection, with additional primary windings to shift the flux waves 45 degrees, could be employed to give the other group of secondary outputs. Such an arrangement would have the disadvantage of requiring many more primary turns on one pair of doublers than on the other pair, with resultant maldistribution of copper losses. The solution of shifting the flux



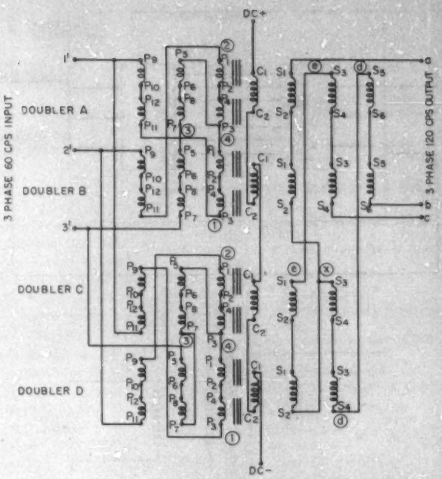


Fig. 5. Three-phase doubler connection

Fig. 4 (left). Duplex doubler connection

A—Doubler A winding arrangement
B—Diagram of connections

waves of each pair of doublers 22½ degrees in opposite directions has therefore been adopted. The necessary arrangement of primary and secondary windings is shown in Fig. 5, and the corresponding vector diagrams are shown in Fig. 6(A). The circled figures and letters shown in Fig. 5 correspond to the figures and letters shown in Fig. 6.

Each of the output vectors of Fig. 6(B) represents the resultant of the emf induced in an individual duplex-connected pair of secondary windings. The method of combination of these output vectors to give a 3-phase output is shown in Fig. 6(C). If the direct control current in either pair of duplex-connected doublers be reversed, then the output vectors from that pair of doublers will be reversed in phase, with consequent phase-sequence reversal of the 3-phase output.

Performance of the 3-Phase Doublers

To test the practicability of the connection shown in Fig. 5, four model doublers were wound on eight small Hypersil C-cores. The maximum output obtained from the model was about 150 watts. The specification of the model doublers is given in Appendix I.

After confirmation of the principles of operation of the 3-phase connection by means of the model doublers, four doublers designed to give a combined output of about 3-kva were built on cores of Stalloy laminations. The specification of the Stalloy doublers is given in Appendix II. In contrast with Hypersil, Stalloy is a nongrain-oriented transformer sheet. The d-c magnetization characteristics for the two materials are shown in Fig. 7.

Open-Circuit Voltage

Variation with control current I_c of open-circuit output voltage per phase E_{oc} and short-circuit current is shown in Fig. 8. The most striking difference in the performance of the two sets of doublers is the rapid fall in open-circuit voltage of the Stalloy doublers after the peak value has been reached, as compared with the slight fall in the case of the Hypersil doublers. Consideration of the characteristics in Fig. 7 shows that for a given core size and configuration the saturated inductance L_s for Stalloy is greater than L_s for Hypersil. For the condition of operation when $I_c > I_k$

$$\theta_s = \mathrm{Sin}^{-1} \frac{2\omega L_s}{\hat{s}_p} (I_c - I_k) \qquad (9)$$

and when $I_c >> I_k$

$$\theta_s \rightarrow \mathrm{Sin}^{-1} \frac{2\omega L_s}{\hat{s}_p} I_c \qquad (10)$$

Thus for two cores of the same size and configuration, a given increase in I_c will cause a greater change in θ_s in the case of Stalloy than in the case of Hypersil. Since the peak value of E_{oc} is reached when $I_c = I_k$, it may be anticipated that E_{oc} will fall away more rapidly from its peak value in the case of Stalloy than in the case of Hypersil.

Short-Circuit Current

The curves of short-circuit current plotted against control current for the two materials, are not dissimilar in shape. The divergence from linearity at high values of control current in the case of Stalloy may be attributed to the same cause as the fall off in open-circuit voltage discussed in the foregoing. It will also

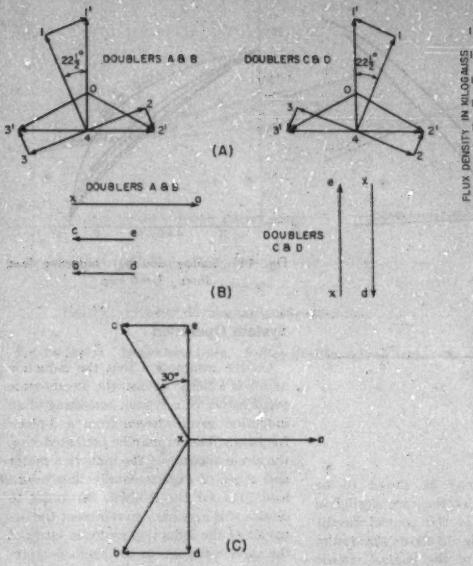


Fig. 6 (left). Three-phase doubler vector diagrams

A—Input vectors
B—Output vectors
C—Vector diagram of combined secondary emf's

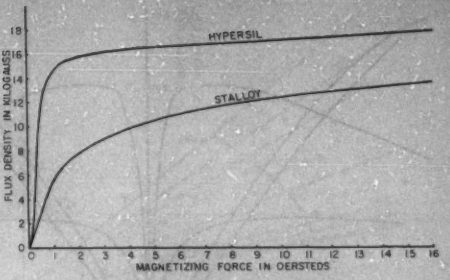


Fig. 7. D-c magnetization characteristics

be noted that while the Stalloy curves pass through the origin, the Hypersil curves do not. The voltage and current intercepts of the Hypersil curves on the vertical axis of Fig. 8 are rms quantities containing no second harmonic and are due to unbalance in the construction of the model doublers.

Internal Secondary Reactance

A quantity of value in forming a physical picture of the operation of the doubler and in choosing a suitable operating range of control current is that of the internal secondary reactance, X_{int}, defined for a particular value of control current as

$$X_{int} = \frac{\text{Open-circuit voltage}}{\text{Short-circuit current}} \qquad (11)$$

From Fig. 8 it will be seen that, disregarding very low values of control current, a curve of X_{int} plotted against control current would resemble a rectangular hyperbola. As a consequence of the variation of internal reactance, the doubler tends to function as a high-impedance current source of low current output for $I_c < I_k$ and increasingly approximates to a voltage source as I_c is increased beyond I_k. As will be seen below, the most advantageous condition of operation of the doubler employed as a power modulator driving an induction motor is reached at values of I_c considerably greater than I_k. At such values of control current, the secondary reactance has fallen to a low value which remains substantially constant for small variations of control current, so that the doubler functions as a voltage source in series with a small reactance.

Load Characteristics

The operating range for the Stalloy doublers may be taken as $0 < I_c < 8$ amperes (amp), since at $I_c = 8$ amp the curve of short-circuit current plotted against control current flattens out markedly. At this value of control current the internal secondary reactance is low and sensibly constant. This estimate of operating range is confirmed by the results of the load tests made on the doublers. Fig. 9 shows the performance of the doublers on resistive load at four different values of control current, and from these load lines it is clear that only a small increase of power output is obtained by increasing the control current from 6 to 8 amp. A control current of 6 amp was selected as the optimum for the drive.

Variation with output current of power output, efficiency, and primary power factor for the resistively loaded double at a control current of 6 amp is shown in Fig. 10. These three curves reach their maxima at approximately the same value of output current, and this coincidence of maxima occurs also at other values of control current. Efficiency and primary power factor fall steadily as control current is increased. At $I_c = 2$ amp efficiency has a maximum value of 0.80 and primary power factor of 0.40. At $I_c = 8$ amp efficiency has a maximum value of 0.64 and primary power factor of 0.24. The efficiency figures given here refer to the operation of the Stalloy doublers on resistive load. With capacitative loading and improvement in design the efficiency can be greatly increased.[1]

Load lines for constant power factor loads, with load impedance varied from open circuit to short circuit are shown in Fig. 11 for a control current of 6 amp. These curves are seen to be similar to those which would be obtained from a source consisting of a voltage with a series linear reactance. At any given control current power output of the doublers, doubler efficiency, and primary power factor all fall rapidly as the doubler load power factor falls from unity. It is thus most desirable to operate the doublers with as high a value of load power factor as possible and therefore to compensate any inductive load by connecting capacitance in parallel with it.

The performance of the Hypersil model doublers on purely capacitative load[4] is shown in Fig. 12. For each value of control current the capacitance across the secondary terminals was increased until ferroresonance developed. The performance of the Stalloy doublers on purely capacitative load was similar to

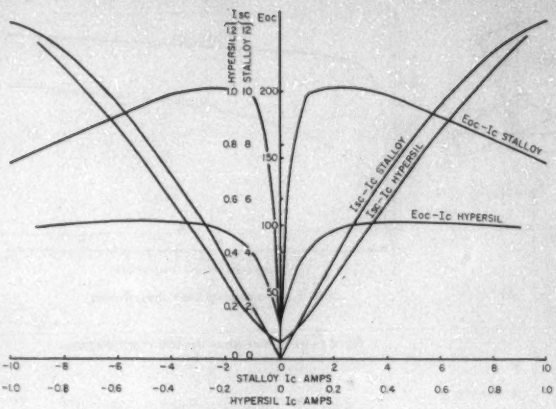


Fig. 8. Doubler characteristics

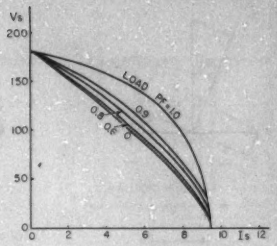


Fig. 11. Stalloy doubler, inductive load lines. $I_c = 6$ amp

that of the Hypersil doublers, but few measurements were made since dangerously high peak voltages developed with the onset of ferroresonance. In practical systems where the possibility of low leading power factor loads exists it would be desirable to connect nonlinear resistors or other voltage limiting devices across the doubler output terminals.

Transient Response

The transient response of the resistively loaded doublers was taken to be essentially the same as that of the doubler control circuit; i.e., the incremental transfer function was assumed to have the form

$$G(S) = \frac{k}{ST_c + 1} \qquad (12)$$

where both the gain factor k, and the time constant T_c are functions of control current. Any other time constants existing in the doubler circuit and the transportation lag which may be shown to be inherent in its operation are negligible in comparison with the control circuit time constant. Fig. 13 shows the results of measurements of the control circuit incremental time constant at various values of control current. The load on the doublers was approximately 75% of the maximum power output at each value of control current. The total resistance of the control circuit was 10 ohms. The dynamic power gain, defined for any value of control current as

$$\text{Dynamic Power Gain} = \frac{\text{Maximum power output}}{\text{Control circuit power input} \times \text{incremental time constant}} \qquad (13)$$

is given below for four values of control current.

I_c	2	4	6	8
Dynamic Power Gain	298	220	153	93

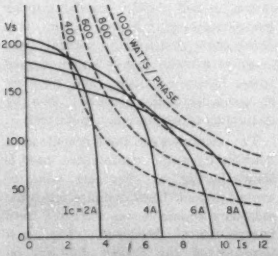


Fig. 9. Stalloy doubler, resistive load lines

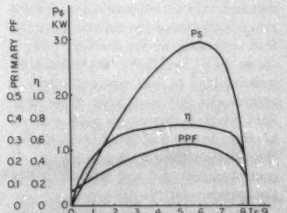


Fig. 10. Stalloy doubler power output, efficiency, and primary power factor. $I_c = 6$ amp

System Operation

On the assumption that the induction motor is a linear device, the steady-state performance of a system consisting of an induction motor driven from a 3-phase frequency doubler may be predicted from the circle diagram of the induction motor and a set of experimentally determined load lines for the doubler. In order to do this it is necessary to represent the behavior of the induction motor in terms of the same variables as are used to represent the behavior of the doubler. The first step in doing this is shown in Fig. 14. Each of the several impedance lines shown in Fig. 14(A) represents a particular condition of operation of the motor for which the corresponding power factor and slip may be determined from the circle diagram. If these impedance lines correspond to power factors for which load lines such as are shown in Fig. 11 have been determined, then the intersections of a load line for a particular power factor with the impedance lines for that power factor will represent two conditions at which the system can operate. A curve may be drawn through all such intersections giving a predicted motor load line upon which the operating point will be decided by the load on the motor.

In order to make the diagram more readable, the constant power factor load lines for the doubler have been omitted from Fig. 14. A set of load lines such as are shown in Fig. 11 for suitable load power factors should be envisaged as superimposed on the impedance lines. The motor load line would then pass through a series of intersections.

The motor load line in Fig. 14(A) has been terminated at the points corresponding to no load operation at the upper end and negative synchronous speed, i.e., slip = 2, at the lower end. For overhauling loads and slips greater than 2, the load

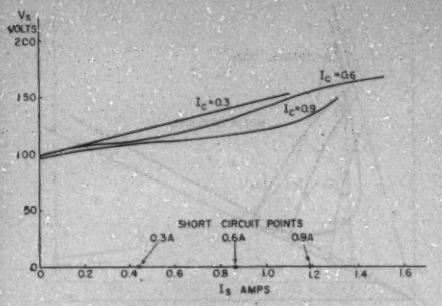


Fig. 12. Hypersil doubler capacitative load lines

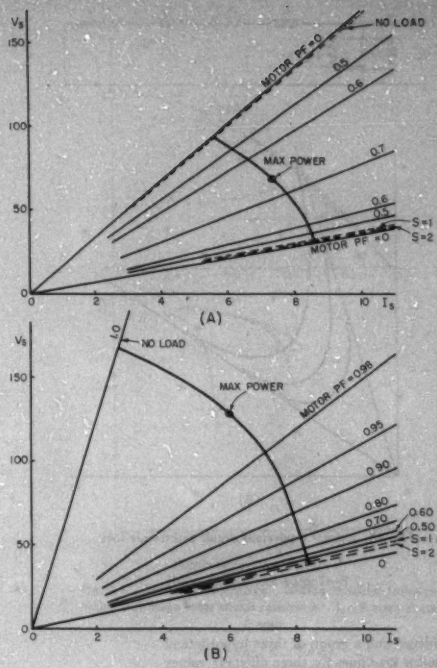


Fig. 14 (right). Motor load lines. Stalloy doubler. I_c = 6 amp

A—Uncompensated
B—Compensated

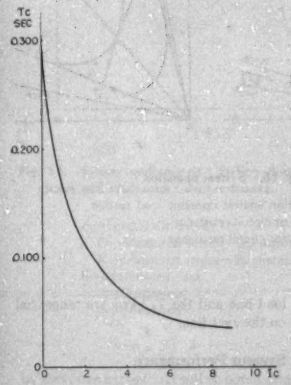


Fig. 13 (left). Time constant of resistively loaded Stalloy doubler

line may be extended as required. The torque-slip curve for the system for the particular value of control current for which the doubler load lines were determined may be predicted from a consideration of a series of points on the motor load line. If the approximation to the linear equivalent circuit for an induction motor shown in Fig. 15(A) be accepted, then it can be shown[6] that the internal torque per phase is given by

$$T = \frac{p}{4\pi f_s} \frac{V_s^2 \frac{sr_3'}{}}{(r_1 s + r_3')^2 + s^2(X_1 + X_3')^2} \qquad (14)$$

where T is expressed in newton meters. For any point on the motor load line, V_s may be read off directly and s may be determined by reference to the motor circle diagram. T may thus be computed and a point on the torque-slip curve fixed.

Capacitance Compensation

The effect on the motor load line of capacitance connected across the doubler output in parallel with the motor is shown in Fig. 14(B). The amount of capacitance employed in order to obtain the result shown in Fig. 14(B) is that required to bring the power factor of the compensated motor to unity on no load. This insures that at any load up to full load the compensated motor operates at a high lagging power factor, and while the doublers function efficiently, the possibility of ferroresonance is excluded. The motor impedance at low slip is now greatly increased, and the impedance at high values of slip is also increased somewhat. The motor load line is thus obtained from the intersections of a much steeper set of impedance lines with a set of constant power factor doubler load lines extending up to unity load power factor. The motor load line is therefore raised and extended as shown. The operating point at which maximum power is delivered to the motor is shown on both of the motor load lines in Fig. 14. For the uncompensated motor a power of 1.04 kw can be delivered by the doublers; for the compensated motor, 2.42 kw.

Operation on Constant-Torque Loads

The second step in representing the behavior of the motor to the same coordinate axes as are used for the doubler is illustrated in Fig. 15(B). For any given torque a series of values of V_s may be determined for a chosen series of values of slip by means of equation 14. If these values of V_s are marked off on the impedance lines corresponding to the chosen values of slip and a line is drawn through the points so obtained, a curve results which is the locus of operation of the motor at the given torque. Three such loci are illustrated in Fig. 15(B). The loci are necessarily asymptotic to the impedance lines for $s = 0$ and $s = \infty$.

The behavior of the system operating against a constant load torque may now be explained in terms of the diagram shown in Fig. 16(A). This shows the torque locus for the torque to be con-

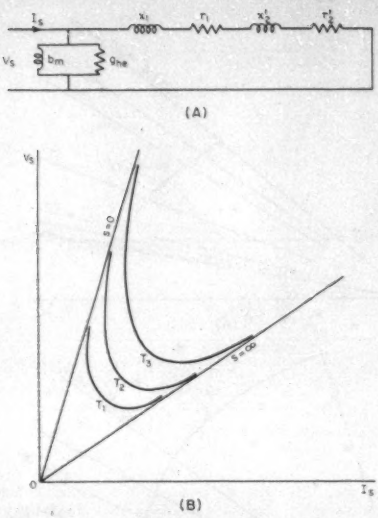


Fig. 15. Motor equivalent circuit and torque loci

A—Motor equivalent circuit
B—Torque loci (hypothetical)

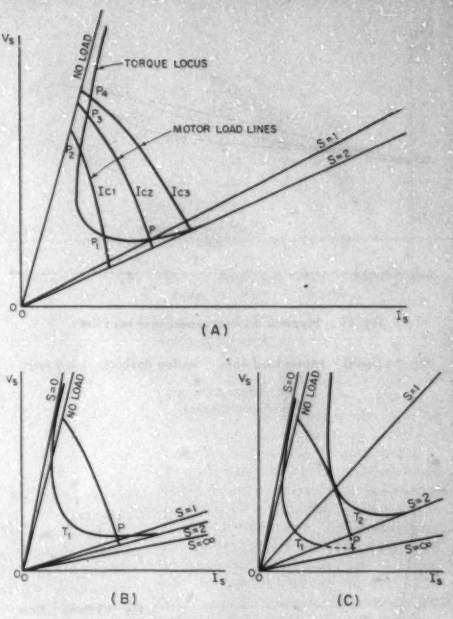


Fig. 16. System operation

A—Operation against constant load torque
B—Low-rotor circuit resistance
C—High-rotor circuit resistance

sidered and a group of three hypothetical motor load lines for three different values of doubler control current. The torque locus, like the load lines, is terminated at the impedance line corresponding to $s=2$. The condition for starting is that the intersection of the torque locus and the motor load line for the control current applied shall lie on or below the impedance line corresponding to $s=1$. This condition is fulfilled at point P for control current I_{c2} and higher values. It is not fulfilled for control current I_{c1} since at point P_1, the motor current is too low to develop a torque equal to that represented by the locus in the diagram. If control current I_{c3} were applied, the motor would accelerate until operating point P_3 was reached. If desired, the control current could then be reduced, and the motor would continue to run. The motor would pull out when the control current had been so far reduced that the motor load line became tangential to the torque locus. If the current were increased to I_{c4}, the system would operate at point P_4.

Effect of Rotor Resistance

Fig. 16 illustrates the effect of additional rotor circuit resistance on the operation of the system. Fig. 16(B) shows a condition in which the motor would be unable to start. The addition of rotor circuit resistance increases the impedance of the motor for all values of slip other than $s=0$ and $s=\infty$ so that the impedance lines corresponding to any given slips increase their slopes within the fixed framework formed by the impedance lines corresponding to $s=0$ and $s=\infty$ as rotor resistance is added. Increase of rotor circuit resistance also has an effect on the torque locus, causing it to approach more closely to the origin of the co-ordinate axes. The effect of a large amount of additional rotor circuit resistance is shown in Fig. 16(C). The intersection of the torque locus for torque T_3 and the motor load line now falls so far below the impedance line for $s=1$ that it is cut off by $s=2$ line. The motor will therefore start. The rotor circuit resistance in Fig. 16(C) has been so chosen that the maximum torque T_3 exerted by the motor occurs at standstill. This is shown by the fact that the motor load line and the T_3 locus are tangential on the $s=1$ line.

System Performance

The motor used in the tests of system performance was rated at 1.5 hp at 115 volts line-to-line with a synchronous speed of 1,200 rpm. When driven from the 120 cycles per second supply provided by the frequency doublers its rating was assumed to be 3.0 hp at 230 volts line-to-line with a synchronous speed of 2400 rpm. Its full load torque for either condition of supply was thus slightly less than 7 foot-pounds.

In Fig. 17(A) are shown theoretical torque-slip curves for the uncompensated and compensated motor with no additional rotor circuit resistance. Measured values are also shown for both conditions. The amount of capacitance used was that required to balance out the reactive component of the motor no-load current at rated motor voltage for the 120 cycles per second supply. The benefit of capacitance compensation is clearly illustrated

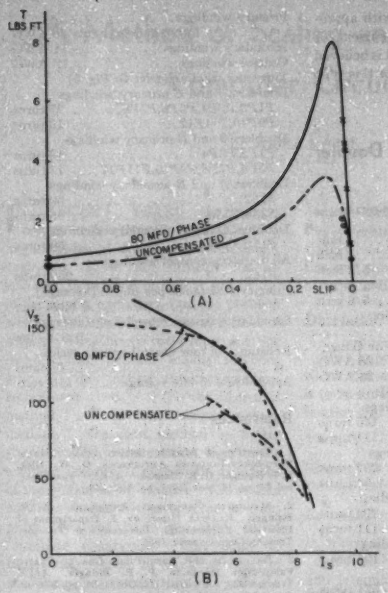


Fig. 17. System performance. Stalloy doubler supplying motor with small rotor circuit resistance. $I_s = 6$ amp

A—Torque slip curves
⊙—Measured values—uncompensated
X—Measured values—80 microfarads per phase
B—Motor load lines
———Theoretical
- - - Measured

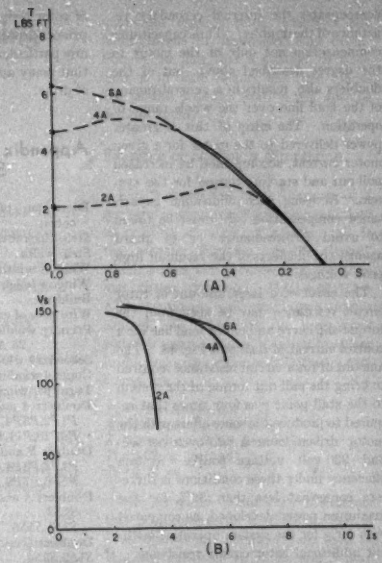


Fig. 18. System performance. Stalloy doubler supplying motor with large rotor circuit resistance. $I_s = 2$ amp, 4 amp, 6 amp

A—Torque slip curves
B—Motor load lines

by these two curves. It will be noted that the pull-out torque for the compensated motor is only slightly greater than the full load torque. In the particular system illustrated, the pull-out torque was determined by the capacity of the doublers, not by the characteristics of the motor. It has been pointed out that the doublers could theoretically be expected to deliver a maximum power of 2.42 kw or 3.26 hp to the motor, and this figure establishes the pull-out torque at 7.60 pounds-feet for a pull-out slip value of 0.06. This corresponds reasonably well with the pull-out torque of 8.0 pounds-feet shown in Fig. 17(A). If a high pull-out torque were desired, then larger doublers would have to be provided. The starting torques in Fig. 17(A) are very low. In practice it would be expedient to use either a high starting torque squirrel cage motor or to include some additional resistance in the rotor circuit of a wound-rotor motor.

In Fig. 17(B) the theoretical and measured motor load lines for the uncompensated and compensated motor are shown. The difference between the theoretical and measured load line in each case is due to the nonlinearity of the motor. For the compensated motor, operating at the rated terminal voltage of 133 volts per phase, the theoretical and measured load lines may be expected to lie close to each other. If the doublers are properly designed, they will deliver full load power to the motor at the rated voltage. At lighter loads, the motor terminal voltage will rise, the motor will saturate, and the compensated motor will operate at a lagging power factor, with consequent reduction of doubler output voltage below that theoretically predicted. Also at very low values of applied voltage, such as the doublers supply on starting the motor, the magnetizing reactance will be substantially higher than at rated voltage, so that while the current drawn by the motor is less than that theoretically anticipated, the power factor of operation will be better than that anticipated. In practice therefore motor current at starting will be lower than the theoretical value while motor terminal voltage will be higher. The difference between theoretical and measured load lines results from a change in motor power factor from that of the linear model. Predictions of power transferred are considerably more accurate, as exemplified by the curves of Fig. 17(A).

Capacitance compensation is of course not limited to the amount used in the tests described in the foregoing. Fig. 12 has shown the type of load line obtained from the doublers with a purely capacitative load. From this it will be seen that, if the capacitance be considered as part of the source, and not as part of the load, then the open circuit voltage of the source is considerably increased by the addition of capacitance which, in effect,

compensates the internal secondary reactance of the doubler. Thus capacitance compensation not only of the motor to the degree described above, but of the doublers also, results in a general raising of the load line over the whole range of operation. The effect of this is greater power delivered to the motor for a given motor current, accompanied by increased pull out and starting torque for the system. In using such additional capacitance compensation care must be taken to avoid ferroresonance or to guard against the ill effects of the resultant high peak voltages.

The effect of a large amount of rotor circuit resistance may be seen from the torque-slip curve and motor load line for a control current of 6 amp in Fig. 18. The amount of rotor circuit resistance required to bring the pull-out torque of the system to the stall point was four times that required to produce the same effect with the motor driven from a 120-cycle-per-second 230-volt voltage source. System efficiency under these conditions is therefore somewhat less than 30% for the maximum power developed, as compared with 50% for the system operating without additional rotor circuit resistance.

The effect of variation of doubler control current is also shown in Fig. 18. At low values of slip, there is little variation in slip with control current at a given torque. Increase of control current increases the power output of the doublers and raises the pull out torque and stalled torque of the system.

Conclusions

The system described above provides a robust and maintenance-free control. Due to the high frequency employed the power obtained from a given size of machine is doubled and a high-speed drive direct from the motor shaft is available. The system may be made inherently torque limiting at any value of torque less than the pull-out torque of the motor. Speed control over the increased speed range may be obtained by variation of rotor resistance combined with appropriate variation of the control current for any particular class of load. It is believed that many applications exist for this type of drive.

Appendix I. Hypersil Doubler Specification

Dimension of Hypersil, single-phase, C cores:
Strip thickness ..... 4 mils
Strip width ..... 1 inch
Window width ..... 5/8 inch
Window height ..... $1^9/_{16}$ inches
Build up ..... 5/8 inch
Wire sizes of enamelled copper wire:
Primary windings ..... 25 AWG (American Wire Gauge)
Secondary windings ..... 28 AWG
Control windings ..... 28 AWG
Turns per winding (refer to Fig. 5):
Doublers *A* and *C* primary windings
P1P2,P3P4,P5P6,P7P8 ..... 155 turns
P9P10,P11,P12 ..... 111 turns
Doubles *B* and *D* primary windings
P1P2,P3P4 ..... 267 turns
P5P6,P7P8,P9P10,P11P12 ..... 64 turns
Doublers *A* and *B* secondary windings
S1S2 ..... 222 turns
S3S4,S5S6 ..... 111 turns
Doublers *C* and *D* secondary windings
S1S2,S3S4 ..... 192 turns
All doublers control windings
C1C2 ..... 574 turns
Primary resistance per phase (average) ..... 3.60 ohms
Secondary resistance per phase (average) ..... 19.0 ohms
Resistance of four control coils in series ..... 54.4 ohms
Applied line-to-line voltage = 95 volts

Appendix II. Stalloy Doubler Specification

Dimensions of laminated Stalloy H cores:
Lamination thickness ..... 14 mils
Outer leg and yoke width ..... $1^1/_2$ inch
Center leg width ..... 3 inches
Window width ..... $1^3/_4$ inches
Window height ..... 8 inches
Depth of stack ..... $3^1/_2$ inches
Wire sizes of enamelled copper wire for primary and control windings, double cotton covered wire for secondary windings:
Primary windings ..... Two parallel strands 9 AWG
Secondary windings ..... 11 AWG
Control windings ..... 12 AWG
Turns per winding (refer to Fig. 5)
Doublers *A* and *C* primary windings
P1P2,P3P4,P5P6,P7P8 ..... 27 turns
P9P10,P11P12 ..... 19 turns
Doublers *B* and *D* primary windings
P1P2,P3P4 ..... 46 turns
P5P6,P7P8,P9P10,P11P12 ..... 11 turns
Doublers *A* and *B* secondary windings
S1S2 ..... 69 turns
S3S4,S5S6 ..... 35 turns
Doublers *C* and *D* secondary windings
S1S2,S3S4 ..... 59 turns
All doublers, control windings
C1C2 ..... 250 turns
Primary resistance per phase (average) ..... 0.0224 ohm
Secondary resistance per phase (average) ..... 0.316 ohm
Resistance of four control coils in series ..... 2.23 ohms
Applied line-to-line voltage ..... 115 volts

References

1. Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers. G. W. Dick, P. P. Biringer, G. R. Slemon. *AIEE Transactions*, vol. 76, pt. II, Nov. 1957, pp. 497–503.

2. Magnetic Frequency Doublers. P. P. Biringer. *Research Report no. 5*, Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada, 1956.

3. Design of the Resistively Loaded Static Frequency Doubler. P. P. Biringer. *AIEE Transactions*, vol. 75, pt. III, Oct. 1956, pp. 834–39.

4. The Magnetic Frequency Doubler with Capacitive Load. P. P. Biringer. *Research Report no. 7*, Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada, 1956

5. Principles of Alternating Current Machinery (book). R. R. Lawrence, H. E. Richards, McGraw-Hill Book Company, Inc., New York, N. Y., 1953, p. 403.

Discussion

J. T. R. Wilson (Aerojet General Corporation, Sacramento, Calif.): This is an outstanding development and it is presented in an easily understood manner.

A. Straughen: Mr. Wilson's comment is appreciated. Investigations into the design of machines best suited to this type of drive and their employment in closed loop speed control systems are in progress.

A Method of Scaling and Checking Computer Circuits

L. JUBIN LANE
ASSOCIATE MEMBER AIEE

IN RECENT YEARS the analog computer has become available to a large number of design and development engineers. It is not difficult for an engineer to become proficient in the use of the analog computer, but one problem which confronts all who use it is that of scaling: How many volts represent one inch? Three popular scaling methods are: (1) the "actual equation-machine equation method,"[1-4] (2) the "per-unit method," and (3) the "direct analogy method." The first method is better suited to the mathematician than to the engineer because it deals with equations of the system being studied. The system equations are rewritten in terms of computer voltages and are then called "machine equations." This tends to cause the engineer to think in terms of the computer rather than to retain a physical picture of the actual equipment he is developing. The other two methods are better in this respect since the thinking tends to be more in terms of actual physical variables. However, when using either of these two methods, many engineers feel that the tie between actual and computer variables is not firm or concrete and that the probability of making a mistake is high enough to introduce uncertainty, especially when integrators or differentiators are involved.

The scaling method presented here has the advantage that it does provide a firm tie between actual and computer variables (by using an equation for scaling) while encouraging the engineer to think in terms of variables and components in the actual equipment. A second advantage is that the scaling equation provides an inherent alarm if a mistake is made. A third advantage is that loop gains may be evaluated to provide an independent check on results.

The same scaling equation is used for all three types of computer problems, namely, simulation, substitution (using the computer as a component of a system), or equation solution. Simulation and substitution are similar and will be explained in the following examples.

Paper 58-42, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 17, 1957; made available for printing November 6, 1957.

L. JUBIN LANE is with the General Electric Company, Waynesboro, Va.

Derivation of Scaling Formula

Fig. 1 presents the general case in which the actual transfer function, β_{xy}, is to be simulated on the computer. The relationship between Y and X is

$$Y=\beta_{xy}X \tag{1}$$

Stage N is to be used to simulate β_{xy} through the analogous relationship

$$E_y=\phi_{xy}E_x \tag{2}$$

in which ϕ_{xy} is the transfer function or "operation" of the computer amplifier and is equal to $Z_1(p)/Z_2(p)$. Although E_x and E_y are both voltages, X and Y are not necessarily voltages, nor are they necessarily the same type of physical quantity. However, it is necessary that E_x and E_y be proportional to X and Y respectively. Consequently, the relationships

$$E_x a_x=X \tag{3}$$

and

$$E_y a_y=Y \tag{4}$$

must be satisfied. The constants a_x and a_y are scale factors and have both numerical values and physical dimensions. The difficulty in scaling is in determining the computer stage gain so that the desired scale factors are obtained. (The converse scaling difficulty is in determining the values of scale factors if the gain is determined first.) Therefore, an equation relating ϕ_{xy}, β_{xy}, a_x, and a_y is desired. This equation is obtained by dividing equation 4 by equation 3 to yield

$$\frac{E_y}{E_x}\frac{a_y}{a_x}=\frac{Y}{X} \tag{5}$$

and then combining equation 1, equation 2, and equation 5 to yield

$$\phi_{xy}=\beta_{xy}\frac{a_x}{a_y} \tag{6}$$

This is the scaling equation. Dimensions should be carried along with numerical quantities since a result which is not homogeneous in dimensions calls attention to the fact that a mistake in stage gain or in scale factors has been made. Use of this scaling equation will now be demonstrated.

Example of Scaling a Simulation

Fig. 2 shows the block diagram of a servomechanism which follows a template to position a cutter on a lathe. K_1 is the transfer function of the "head" which senses the difference in template and cutter positions. $M/(1+pT_1)(1+pT_2)$ represents an amplidyne motor control. K_3 represents the lead screw which translates motor rotation to cutter lineal velocity. Values of these constants are

$K_1=1.2$ volts per mil
$M=600$ rpm per 100 volts
$K_3=1$ inch per 5 rev (revolutions)
$T_1=0.02$ sec (second)
$T_2=0.02$ sec

The computer circuit used to simulate this position control is given in Fig. 3. Five operational amplifiers are used so that the variables y_t, y_c, m, and S may be recorded. A knowledge of the maximum expected values of the variables enables selection of scale factor values such that no computer amplifier will be overdriven. Table I gives these maximum expected values and the scale factor selected. (Here it is assumed that the computer amplifier output can swing ± 100 volts. A computer volt is designated a unit.)

The operation from point 6 to point 7 (that is, ϕ_{6-7}) will now be evaluated. Inserting the appropriate subscripts into equation 6 gives

$$\phi_{6-7}=\beta_{6-7}\frac{a_6}{a_7} \tag{7}$$

Table I gives values of the scale factors, but the value of β_{6-7} must be taken from the system being simulated. Since ϕ_{6-7} simulates the actual transfer function from S to y_c,

$$\beta_{6-7}=\frac{y_c}{S} \tag{8}$$

Fig. 2 shows that this transfer function is

$$\frac{y_c}{S}=\frac{K_3}{p} \tag{9}$$

and since $K_3=1$ inch per 5 rev,

$$\beta_{6-7}=\frac{1 \text{ inch}}{p5 \text{ rev}} \tag{10}$$

Substituting into equation 7 from equa-

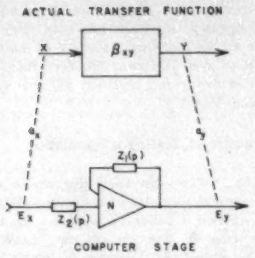


Fig. 1. Analog representation of a transfer function

tion 10 and Table I yields

$$\phi_{8-7}=\frac{1}{p5}\frac{\text{inch}}{\text{rev}}\times\frac{6\text{ rpm/unit}}{10\text{ mils/unit}} \tag{11}$$

$$=\frac{1}{p5}\frac{6}{10}\frac{\text{inch}}{\text{rev}}\times\frac{\text{rev}}{\text{min}}\frac{1}{\text{mils}}\left[\frac{1{,}000\text{ mils}}{1\text{ inch}}\times\frac{1\text{ minute}}{60\text{ sec}}\right] \tag{12}$$

in which the bracketed terms are conversion factors equal to unity. Canceling dimensions in equation 12 yields

$$\phi_{8-7}=\frac{2}{p\text{ sec}} \tag{13}$$

From Fig. 3,

$$\phi_{8-7}=\frac{1}{pR_8C_8} \tag{14}$$

(it is not necessary to include the minus sign introduced by the operational amplifier), and by combination of equation 13 and equation 14

$$R_8C_8=\frac{1}{2}\text{ sec} \tag{15}$$

Hence, a 1/2-sec rate of integration ($R_8=0.5$ megohm and $C=1.0$ microfarad, for instance) gives stage D the value of operation necessary to simulate the transfer function y_c/S while the two scale factors a_6 and a_7 apply.

To clarify, this procedure is

1. To select a segment of the computer circuit to be scaled, ϕ_{xy}.
2. Identify the corresponding segment of the actual system, β_{xy}.
3. Select scale factors consistent with the allowable computer amplifier swing.
4. Apply equation 6.

This method will now be applied in scaling the remainder of the computer circuit.

Stage F is merely for sign changing, so

$$\phi_{7-6}=\frac{R_{10}}{R_9}=1 \tag{16}$$

and substitution into equation 6 yields

$$1=\beta_{7-6}\frac{a_7}{a_6} \tag{17}$$

Since β_{7-6} is simply y_c to y_c, $\beta_{7-6}=$ unity and $a_6=a_7=10$ mils per unit.

Stage A has two inputs, but its two operations can be evaluated by superposition since the circuit is linear to both inputs. The operation is ϕ_{2-1}, and β_{2-1} is from y_c to m. Fig. 2 shows that

$$m=y_t-y_c \tag{18}$$

which means that the transfer function from either y_t or y_t to m is unity. Hence

$$\phi_{2-1}=1\frac{10\text{ mils/unit}}{1\text{ mil/unit}}=10 \tag{19}$$

and $R_2=R_3/10$ must be used. In addition $R_1=R_5=R_3/10$ must be used since $a_2=a_3$ and since $\beta_{2-1}=\beta_{6-1}$.

In the actual system m and e_p are proportional, so they can be represented by the same voltage E_1. Equation 6 is applicable here also. From Fig. 2,

$$\beta_{1-3}=\frac{e_p}{m}=K_1 \tag{20}$$

and since $\phi_{1-3}=1$,

$$a_3=\frac{K_1}{1}a_1 \tag{21}$$

$$=\frac{1.2\text{ volts}}{\text{mil}}\frac{1\text{ mil}}{\text{unit}} \tag{22}$$

$$=1.2\text{ volts/unit} \tag{23}$$

Since E_3 is not a counterpart of anything in the actual system, stages B and C may be scaled as one operation, ϕ_{1-5}.

Fig. 3 shows that β_{1-5} is a transfer function from m to S, and Fig. 2 shows that

$$\frac{S}{m}=\frac{K_1M}{(1+pT_1)(1+pT_2)}=\beta_{1-5} \tag{24}$$

Substituting into equation 6 yields

$$\phi_{1-5}=\frac{1.2\text{ volts}}{\text{mil}}\frac{600\text{ rpm}}{100\text{ volts}}\frac{1\text{ mil}}{\text{unit}}\frac{}{(1+pT_1)\,(1+pT_2)\,6\text{ rpm/unit}} \tag{25}$$

$$=\frac{1.2}{(1+pT_1)(1+pT_2)}\text{ (dimensionless)} \tag{26}$$

Fig. 2. Integrating servomechanism to be simulated

Table I. Simulation Scale Factors

Variable	Maximum Value Expected	Scale Factor	Values of Scale Factor
m	75 mils*	a_1	1 mil/unit
y_t	1,000 mils	a_2	10 mils/unit
y_c	1,000 mils	a_3 and a_4	10 mils/unit
e_p	90 volts	a_5	1.2 volt/unit
t	600 rpm	a_6	6 rpm/unit

* 1 mil = 0.001 inch.

From Fig. 3 it is seen that

$$\phi_{1-6}=\frac{R_4}{R_6(1+pR_4C_4)}\frac{R_7}{R_6(1+pR_7C_7)} \tag{27}$$

which with equation 26 shows that the simulation is correct if

$$R_4C_4=T_1 \tag{28}$$

$$R_7C_7=T_2 \tag{29}$$

and

$$\frac{R_4R_7}{R_6R_6}=1.2 \tag{30}$$

Actually, it is not necessary to include the "$1+pT$" terms in equation 6, since scaling steady-state values is sufficient if the proper time constants are used.

Now that the entire loop has been scaled, it should be checked by comparing the open loop gain of the computer circuit (Comp GH) with that of the actual system ($T_{\text{inst}}\,GH$). Two systems behave identically if their GH's are identical. It is possible for the GH of the computer loop to be equal to that of the actual system since GH is always a dimensionless number. The GH of the actual system,

$$T_{\text{inst}}\,GH=\frac{K_1MK_6}{p(1+pT_1)(1+pT_2)} \tag{31}$$

$$=\frac{24}{p(1+pT_1)(1+pT_2)\text{ sec}} \tag{32}$$

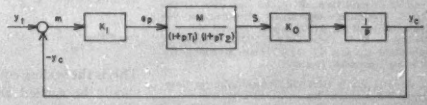
The GH of the computer circuit is

$$\text{Comp }GH=\phi_{1-6}\,\phi_{6-7}\,\phi_{7-5}\,\phi_{5-1} \tag{33}$$

$$=\frac{1.2}{(1+p0.02)(1+p0.02)}\frac{2}{p\text{ sec}}(1)\ 10 \tag{34}$$

$$=\frac{24}{p(1+p0.02)(1+p0.02)\text{ sec}} \tag{35}$$

which is correct and verifies the scaling done by use of equation 6.



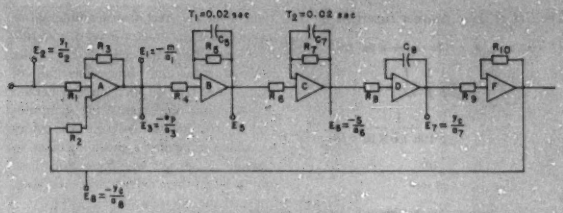


Fig. 3. Computer circuit for simulation of the integrating servomechanism

Example of Scaling an Equation

Next, scaling of a computer circuit for solving a differential equation will be illustrated. The voltage equation of a series inductance-resistance-capacitance circuit

$$Lpi+Ri+\frac{1i}{pC}=e \tag{36}$$

will be taken as an example. It is customary to solve the differential equation for its highest derivative and to employ integration instead of differentiation. Thus, equation 36 is written

$$pi=\frac{e}{L}-\frac{Ri}{L}-\frac{1i}{pLC} \tag{37}$$

and the computer circuit in Fig. 4 may be employed. In this case

L=10 h (henry)
$C=1.0\times10^{-4}$ farads
R=30 ohms
e=300 sin $(2\pi ft)$ volts
1 cps$\leq f\leq$20 cps

Table II gives maximum expected values of the variables and the scale factors selected.

Equation 6 can be employed in scaling this computer circuit because for every computer operation from one labeled voltage to another, there is a corresponding mathematical operation. This operation is the β_{xy} in Fig. 1.

For instance, the operation of stage 3 is ϕ_{21-22} which corresponds to a transfer from pi to i. Since this transfer is an integration, β_{21-22} is simply $1/p$. Thus,

$$\phi_{21-22}=\frac{1}{pR_dC_d}=\beta_{21-22}\frac{a_{21}}{a_{22}} \tag{38}$$

$$\frac{1}{pR_dC_d}=\frac{1}{p}\frac{\frac{5\text{ amp}}{\text{sec/unit}}}{0.1\text{ amp/unit}} \tag{39}$$

in which ampere is abbreviated "amp."

$$R_dC_d=0.02\text{ sec} \tag{40}$$

Hence, the operation is correct and the scale factors are those given in Table II if the resistance-capacitance product used in stage 3 is 0.02 sec.

To ascertain what mathematical operation ϕ_{23-24} represents, superposition is applied. By imagining E_{23} and E_{24} to be zero, it is seen that the e/L and the i/pCL terms in E_{24} are both zero and that the Ri/L term is a result of the E_{23} input. Consequently, β_{23-24} is a transfer function from i to Ri/L, namely, R/L.

Since ϕ_{23-24} is R_d/R_f, applying equation 6 yields

$$\phi_{23-24}=\frac{R_d}{R_f}=\beta_{23-24}\frac{a_{23}}{a_{24}} \tag{41}$$

$$\frac{R_d}{R_f}=\frac{R}{L}\frac{0.1\text{ amp/unit}}{\frac{5\text{ amp}}{\text{sec/unit}}} \tag{42}$$

$$\frac{R_d}{R_f}=\frac{30\text{ ohms}}{10\text{ h}}\frac{0.1\text{ sec}}{5} \tag{43}$$

Since the henry is equivalent to a volt-sec/amp, equation 43 reduces to

$$\frac{R_d}{R_f}=0.06\text{ (dimensionless)} \tag{44}$$

and letting R_d be 0.1 megohm requires that R_f be 1.67 megohms.

By a similar procedure stage 5 is found to require an R_bC_b product of 0.05 sec, R_b is found to be 0.10 megohm, and R_g is found to be 1 megohm.

Table II. Equation Scale Factors

Variable	Maximum Expected Value	Scale Factor	Value of Scale Factor
pi	314 amp/sec	a_{21} and a_{22}	$\frac{5\text{ amp}}{\text{sec/unit}}$
i	10 amp	a_{23}	0.1 amp/unit
i/p	0.318 coulomb	a_{24}	0.005 coulomb/unit
e	300 volts	a_{25}	5 volts/unit

The next step to be illustrated is the method of checking loop gains. The computer circuit open-loop gain, Comp GH, must be equal to the open-loop gain of the mathematical transfer functions, $TfGH$. From Fig. 4, the computer circuit is seen to have two closed loops (taken in the directions signals travel). The loop through points 21, 22, 24, 26, 21 has an open loop gain of

$$\text{Comp } GH_{21,\,22,\,24,\,26,\,21}=\frac{1}{p0.02\text{ sec}}\frac{0.1}{1.67} \tag{45}$$

$$=\frac{3}{p\text{ sec}} \tag{46}$$

The mathematical transfer functions in this path are

pi to i $\left(\beta_{21-22}=\frac{1}{p}\right)$

and

i to $\frac{Ri}{L}$ $\left(\beta_{23-24}=\frac{R}{L}\right)$

The GH of the transfer functions is

$$TfGH_{21,\,22,\,24,\,26,\,21}=\beta_{21-22}\beta_{23-24} \tag{47}$$

$$=\frac{1}{p}\frac{R}{L}=\frac{3}{p\text{ sec}} \tag{48}$$

If a mistake had been made in using equation 6 for scaling, Comp GH and $TfGH$ would have been unequal, calling attention to the fact.

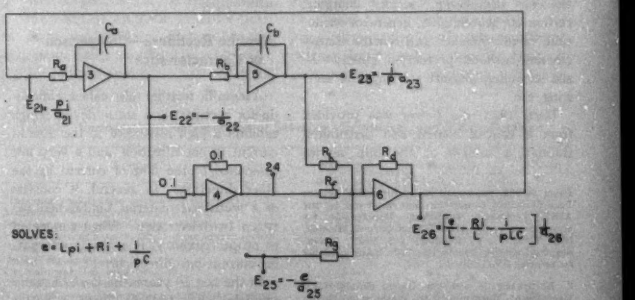


Fig. 4. Computer circuit for solving a second-order differential equation

The other loop will be checked also. From Fig. 4,

$$\text{Comp } GH_{21,\ 22,\ 23,\ 24,\ 21} = \frac{1}{p0.02 \text{ sec}} \times \frac{1}{p0.05 \text{ sec}} \frac{0.1}{0.10} \quad (49)$$

$$= \frac{10^3}{p^3 \text{ sec}^3} \quad (50)$$

The mathematical transfer functions in this loop are

pi to i $\left(\beta_{21-22} = \frac{1}{p}\right)$

i to $\frac{i}{p}$ $\left(\beta_{22-23} = \frac{1}{p}\right)$

$\frac{i}{p}$ to $\frac{i}{LCp}$ $\left(\beta_{23-24} = \frac{1}{LC}\right)$

The GH of these transfer functions is

$$TfGH_{21,\ 22,\ 23,\ 24,\ 21} = \beta_{21-22}\,\beta_{22-23}\,\beta_{23-24} \quad (51)$$

$$= \frac{1}{p} \frac{1}{p} \frac{1}{LC} \quad (52)$$

$$= \frac{1}{p^2 10\text{h } 1.0 \times 10^{-4} fd} \quad (53)$$

$$= \frac{10^3}{p^2 \text{ sec}^2} \quad (54)$$

Again the loop gains evaluated by two independent methods provide a valuable check against mistakes.

Summary

Equation 6 provides a dependable formula for scaling a computer circuit. In evaluating a computer stage operation, if an incorrect transfer function or an incorrect scale factor is used, equation 6 is usually rendered dimensionally heterogeneous and dimensions fail to cancel properly. In addition, the method of evaluating and comparing computer and transfer function loop gains provides an added safeguard against mistakes in scaling. Engineers who have used these methods find them particularly satisfying.

References

1. Servomechanisms and Regulating System Design (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc., New York, N. Y., vol. 1, 1951.

2. Electronic Analog Computers (book), G. A. Korn, T. M. Korn. McGraw-Hill Book Company, Inc., New York, N. Y., 1952, pp. 22–115.

3. Analog Computer Techniques (book), C. L. Johnson. *Ibid.*, 1956, pp. 20–44.

Metallic Rectifiers for Shipboard Electric Systems

C. L. STRAUB
ASSOCIATE MEMBER AIEE

H. G. WIEST
MEMBER AIEE

PRIOR TO 1932, all Navy ships had d-c electric plants for supplying power to electric auxiliaries and lighting systems. As the size of electric plants grew, it was recognized that the adoption of a-c systems for shipboard electric plants offered distinct advantages. At the present time, all of the newer Navy shipboard electric generating plants, except for some small ships, are designed to provide 450 volts alternating current. However, some electric auxiliaries used on board ship require direct current for their operation. Direct current is needed for such applications as degaussing, battery charging, carbon-arc searchlights, generator excitation, circuit breaker and reactor control circuits, cathodic protection, guided-missile servicing, aircraft starting and servicing, etc.

Until 1951, d-c power was provided from a central source and distributed through a section of the ship service switchboard. It was provided by using either motor-generator sets or oversized exciters on the a-c generators. Thus, on ships with a-c power plants, a complete d-c power generation and distribution system was provided that was analogous to the a-c power system.

The development of metallic rectifiers has resulted in the applications of rectifiers as d-c power sources on ships with a-c power plants. Rectifiers are now being used as direct current power sources on the Forrestal-class aircraft carriers, guided-missile destroyers, and other ships.

Metallic Rectifiers—Comparison of Characteristics

A metallic rectifier (also called semiconductor rectifier) is a static device which exhibits a high resistance to the flow of current in one direction, and a very low resistance to the flow of current in the other direction. In general, it consists of a rectifying material sandwiched between two electrodes. When connected to proper circuitry, it converts alternating current into direct current.

In the last 10 years, the work of many physicists has brought forth a better understanding of the mechanism of rectification which has resulted in rectifier cells of higher voltage, higher temperature operation, and generally improved performance. The most important metallic rectifying materials at present are selenium, germanium, and silicon.

Selenium rectifiers are unique in that they require forming as a final step in manufacture. This is accomplished by passing a current through the cell for a period of time during which the reverse resistance build up to a maximum value. Conversely, a selenium rectifier tends to lose its high reverse resistance when standing idle for a long period of time. However, it rapidly regains its original reverse resistance when re-energized.

When subjected to small overvoltage surges, the weaker spots in the cell tend to burn out, but these spots are "self-healing," and the performance of the cell is not seriously impaired. Similarly, overcurrent surges cause momentary heating, but the cell has sufficient thermal capacity to absorb these surges without serious damage.

Selenium rectifiers are subject to aging with use, which causes an increase in forward resistance. This shows up as a decrease in output voltage and design provisions (such as transformer taps) must be made to compensate for this effect.

A selenium rectifier is said to have finished its useful life when its forward resistance has doubled. When so applied that the cell temperature is kept below 75 C (degrees centigrade), a life of about 80,000 hours is obtained. Higher operating temperatures shorten the life, so that with a maximum cell temperature of 150 C (ambient 130 C) the life is about

Paper 58-135, recommended by the AIEE Marine Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 16, 1957; made available for printing November 12, 1957.

C. L. Straub is with the United States Navy Department, Bureau of Ships, Washington, D. C.; H. G. Wiest is with the General Electric Company, Lynchburg, Va.

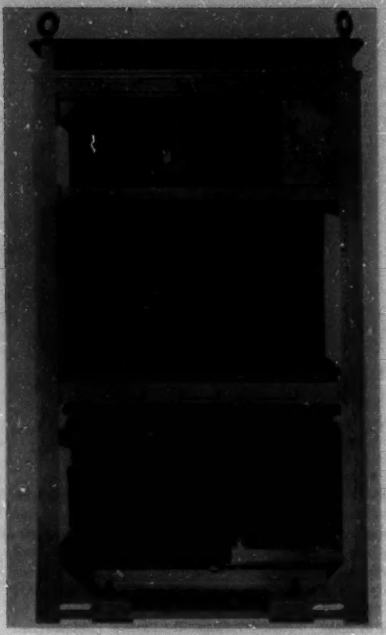


Fig. 1. Degaussing power supply, 25.4 kw (Vickers Electric Division)

Fig. 2. Jet aircraft servicing and starting power supply (General Electric Company)

1,500 hours without derating. For shipboard applications the normal ambient temperature is 50 C maximum.

Modern selenium cells withstand a continuous peak inverse voltage of 63 volts. The efficiency of a selenium transformer-rectifier may be as high as 85%. The losses in the rectifier cells may be handled with convection cooling, but a two- or three-fold reduction in size may be obtained with forced-air cooling, depending on the rating. This is particularly true for rectifiers rated 50 to 100 kw. Due to long experience and mass production methods, selenium rectifiers have generally been found less expensive than germanium or silicon.

Germanium rectifiers are comparatively new with application experience in the power field dating back only about 5 years. If kept below 75 C they are stable and show no aging effects. Because of the high current density at which they operate, they are very small compared to selenium but for the same reason they are destroyed by overvoltage or overcurrent surges. Considerable attention must be given to the cooling of germanium cells because the losses are concentrated in such a small area. The extremely high reverse resistance makes germanium rectifiers very attractive for magnetic amplifier applications. The reverse current is so small that, even though it varies with temperature, it has negligible resetting effect on magnetic amplifier cores in many applications. Currently, the maximum continuous peak inverse voltage for germanium cells is about 100 volts at a cell temperature of 75 C. The efficiency of a germanium transformer-rectifier may be as high as 94%. Continuing research may improve some or all of these limitations.

Silicon rectifiers are quite new, and have properties resembling those of germanium. Their outstanding differences are much higher peak inverse voltage (300 volts), a much higher maximum operating temperature, 200 C, and a maximum current density several times as great as germanium. The maximum efficiency

Table I. Comparison of the Properties of Selenium, Germanium, and Silicon Rectifiers

Property	Selenium	Germanium	Silicon
Maximum peak inverse voltage per cell, volts	63	100	300
Maximum operating cell temperature, C	75*	75	200
Max. d-c output voltage per cell, volts†	54	80	270
Forward current density, amperes per square inch†	1.2	1,000	3,000
Forward voltage drop at above density, volts	1.2	0.7	1.3
Typical regulation, new, per cent	10	5	5
Typical regulation, aged, per cent	15	5	5
Cell failure protection required	no	yes	yes
Typical efficiency, transformer and rectifier, per cent	85	94	94

* For 80,000-hour life. † For 3-phase full-wave bridge circuit, forced air cooled.

Fig. 3. Static excitation system mounted in the air-cooler of an 800-kw turbogenerator set for a nuclear-powered submarine (Westinghouse Electric Corporation)

Fig. 4. End view of 2,000-watt alternator rectifier showing silicon rectifier installation (Leece Neville Company)

is about the same as for germanium. Cooling means are even more important than with germanium due to the even greater concentration of losses. Silicon rectifiers are also well adapted to magnetic amplifier applications, especially where high temperature operation is required.

Table I compares the properties of selenium, germanium, and silicon rectifiers. These values should be considered typical rather than limiting.

It is not always possible to realize all of the values given in the above table in a specific design. The selection of the optimum rectifying medium is a complex problem depending upon the particular application. Most rectifier manufacturers have experienced application engineers to analyze specific requirements and recommend a suitable choice and construction.

Rectifier Applications

Where rectifiers are used, it is the practice to provide a rectifier for each d-c load or auxiliary and install it in the vicinity of the equipment for which it is designed. Some typical applications of metallic rectifier power supplies are discussed in the following.

Degaussing

A degaussing system is installed aboard ship to minimize the possibility of the ship actuating a magnetic firing device such as used in magnetic mines and magnetic torpedoes. This system consists of one or more coils of electrical cable wound in a specific location inside the ship's hull. A direct current is passed through these coils which produces counter magnetic fields that are at all times equal in magnitude but opposite in polarity to the ship's fields. The magnitude and polarity of the current is controlled by auxiliary equipment which is either manually or automatically operated. Degaussing coil power supplies have 440-volt 60-cycle a-c power inputs and d-c outputs. Rectification in the power supply is accomplished by means of selenium rectifiers. The magnitude of the output current is controlled by magnetic amplifiers. A ventilating fan usually provides cooling for the rectifier bank. Degaussing power supplies range from 3 to 40 kw depending on the degaussing coil and size ship. For a recent class of Navy oilers, the ratings of the degaussing power supplies are 7.2 kw, 25.4 kw, and 40 kw for the various degaussing coils. The rectifier power supply for one of the coils is shown in Fig. 1.

Aircraft Servicing and Starting

The primary function of the deck edge low-voltage electric system found on most types of modern aircraft carriers is to provide a readily accessible source of servicing and starting power to jet aircraft at almost all locations on the carrier's flight deck. Servicing power from 25 to 29 volts direct current is used to warm up and check out the aircraft's electronic equipment and other vital equipment before take-off. It is also required for routine test and maintenance of various items of aircraft equipment. Modern aircraft require approximately 1,100 amperes, 33 volts direct current for starting jet engines.

Low-voltage d-c power for aircraft servicing and starting is obtained from specially designed rectifier power supplies. One of the aircraft servicing and starting rectifier power supplies for an aircraft carrier is shown in Fig. 2. These power supplies for an aircraft carrier is shown in Fig. 2. These power supplies are designed to maintain constant voltage at the servicing terminals regardless of the load (from 0 to 600 amperes). Current to jet engine starting motors is held between 1,000 and 1,100 amperes by magnetic amplifier regulators with a finishing voltage of approximately 33 volts direct current.

Static Excitation Systems

Static excitation systems are being furnished for some ships service and emergency generators. Minesweepers, some nuclear powered submarines, and the 400-cycle turbine generator sets on aircraft carriers have static excitation systems. In these systems direct current for generator field excitation is provided by selenium or silicon rectifiers. The rectifiers are either mounted remote from the generator or are mounted on the generators so that they can be cooled with air circulated by the generator fan. A typical static excitation system rectifier as-

Table II. Analysis of the Advantages and Disadvantages of Metallic Rectifiers Versus Motor-Generator Sets

Characteristic	Motor-Generator	Rectifier
Light load operation	B	A
Reliability	B	A
Maintenance	B	A
Size	B	A
Weight	B	A
Cost	B	A
Control	B	B
Overload	B	B
Radio interference	B	A
Quiet operation	B	A
Efficiency	B	A
Output fixed voltage	B	A
Output variable voltage	A	B

A = Superior
B = Suitable

sembly mounted on top of a turbine generator in the air cooler is shown in Fig. 3.

Magnetic Amplifiers

Voltage regulator and various control applications are being accomplished with the use of magnetic amplifiers. Magnetic amplifiers usually employ rectifiers as circuit elements in conjunction with saturable reactors. Rectifiers will actually pass a small (leakage) current in the reverse direction. Leakage current in magnetic amplifiers results in desaturating the core and lowering the gain of the amplifier circuit. Because of this, rectifiers with low leakage currents are required for magnetic amplifiers. In addition to being used for voltage regulation, magnetic amplifiers have found wide application in control circuits, such as nuclear reactor instrumentation.

Electronics

Metallic rectifiers have completely changed the concept of d-c power supply design. Where it had previously been considered impractical to use voltage multiplier circuits because of the complexity in supplying filament power, rectifiers are now in common use. Voltage multipliers using metallic rectifiers eliminate the need for power transformers and rectifier tubes. Because metallic rectifiers are efficient, cool operating, instantaneous in action and allow compact lightweight designs, they are used in all types of electronic equipment.

Cathodic Protection Systems

Shipboard cathodic protection installation supplement the paint coating in preventing corrosion of the underwater hull and its appendages. Protection against corrosion is achieved by eliminating the potential difference between anodic and cathodic areas. A cathodic protection system is a closed-loop system in which a power supply produces a current flow from isolated anodes through the sea water to the ship's hull in a quantity such that the hull potential remains essentially constant at a predetermined optimum value. The Bureau of Ships is obtaining prototype automatically controlled d-c power supplies for use in cathodic protection systems. Metallic rectifiers will be used in these power supplies.

Small Boat Generators

In general, small landing craft and boats are provided with d-c generators belt driven from the propulsion engine. At low propulsion engine speeds the d-c generators are not capable of supplying power. With increased electrical loads and the addition of electronic equipment on small landing craft and boats, it is necessary to supply power over the complete propulsion engine speed range in order to prevent the boat's storage battery from becoming discharged. To meet these requirements, an alternator rectifier was developed for Navy use. The first alternators used selenium rectifiers for converting the 3-phase alternating current to direct current. Propulsion engine intake air was used for cooling the rectifiers. This was a satisfactory electrical installation. However, it required modification of the air intake manifold on the diesel propulsion engine so that the rectifier assembly could be accommodated. To overcome this difficulty, the manufacturer developed a 2,000-watt 28.5-volt d-c alternator rectifier using silicon rectifiers as shown in Fig. 4. Because of their small size, the silicon rectifiers are mounted inside the generator enclosure. This type of rectifier installation permits blast cooling of the rectifiers by means of air supplied by the generator cooling fan. An alternator rectifier of this type was successfully evaluated and is being specified for recent small boat programs.

Battery Charging

Metallic rectifier power supplies are used aboard ship for charging the various storage batteries installed for engine starting, portable flood lanterns, etc. The rectifier power supply is part of the battery charging switchboard which is arranged to charge the various batteries at either local or remote stations. Batteries range from 2 to 24 volts.

Circuit Breaker Control

Metallic rectifiers are used to convert the a-c control source to direct current which is required by the closing coil of electrically operated circuit breakers. Selenium rectifiers are used, although recently the Navy has authorized the use of silicon rectifiers for this purpose. In addition to the above applications, rectifiers are being used in the power supplies for guided-missile servicing and carbon-arc searchlights. One of the largest rectifier installations on board ship is that for the stern anchor winch on a new class of landing ships. This power supply is rated at 100 kw.

Table III. Comparison Between a Motor Generator and a Metallic Rectifier Installation

	Central Motor Generator Installation, Per Cent	Selenium Rectifier Installation, Per Cent
Cost	100	40 to 50
Weight	100	50 to 60
Volume	100	60 to 70

Metallic Rectifiers Versus Motor-Generator Sets

In selecting a d-c power supply, these factors should be considered: application, reliability, maintenance, size, weight, cost, and efficiency. An analysis of the advantages and disadvantages of metallic rectifiers versus motor-generator sets is presented in Table II.

Metallic rectifiers have a decided advantage when operating at light loads because there is no decrease in life and their efficiency is high. Compared to this, motor-generator sets have certain disadvantages. Running motor-generator sets at light loads tends to destroy the film on the commutator which in turn produces brush wear and some brush chatter. This adds to maintenance expense in that the commutator may become rough. Also, the efficiency of motor-generator sets is quite low at light loads. On this basis the rectifier is more suitable for light load applications.

Maintenance time and expense for metallic rectifiers are much less than for rotating equipment since there are no moving parts except, in some cases, a cooling fan. Inherently, rotating machines require more maintenance.

A comparison between a motor-generator and a metallic rectifier installation was made for a recent class of minesweepers and escort vessels in order to show the weight, space requirements, and cost of either installation. For the rectifier installation, power supplies were provided at the individual loads. For the motor-generator installation a central d-c

system was used. Table III shows this comparison.

The savings indicated in Table III result, to a great extent, from the fact that the rectifier installation does not provide for two separate power sources as is done in the case of motor-generator sets. The rectifier, being a static device, has no moving parts which require servicing whereas the motor-generator requires frequent inspection and servicing. To permit this servicing, a stand-by motor-generator set is provided so that a d-c power supply is always available.

On a ship basis, the use of individual metallic rectifiers in lieu of a central motor-generator set for d-c power supplies has resulted in substantial savings in cost, weight, and space as noted in Table III. Also, the use of rectifiers has reduced the cable requirements and the number of bulkhead penetrations since long runs of cables carrying direct current have been omitted. Since rectifiers are dispersed throughout the ship, most of the space and weight savings take place in the generator or engine room where space is at a premium. With the further development and application of silicon and germanium rectifiers, the weight and space requirements will be reduced further.

Another advantage of metallic rectifier power supplies is that they do not require special foundations. Rectifier units can be mounted on the deck or bulkhead without any difficulty whereas motor-generator sets must be provided with foundations which will assure proper alignment of the motor-generator set.

If adjustable-voltage direct current is required over a wide range, the motor-generator set is probably the best conversion means. However, rectifier power supplies can be made to provide adjustable voltage by using saturable reactors or induction voltage regulators in the a-c input voltage.

Metallic rectifiers have the advantage of quiet and vibrationless operation. They are simple and inexpensive to install, operate reliably unattended, and have low maintenance and high momentary load capacity.

Also, the use of metallic rectifiers on board ship reduces the number of spare components required because of their reliability.

Installation and Operating Features

Metallic rectifiers, particularly selenium and germanium, are designed for operation in a well-ventilated location so that the total temperature of the rectifier cell does not exceed the rated value. To operate them at their specified electrical rating and to expect the desired reliability and low maintenance requires that adequate ventilation be provided.

Rectifier power supplies for shipboard applications are normally designed for use in 50 C ambient temperatures. Installing activities have the responsibility of seeing that rectifier power supplies are located in surroundings where the ambient temperature is compatible with the equipment design ambient.

When a location for rectifier power supplies is being chosen, the following factors should be considered in order that the heat from the rectifiers may be dissipated promptly:

1. Equipment that contains metallic rectifiers should be located in the coolest available place close to the power consuming equipment.

2. The location selected should not be downstream from other equipment that dissipates much heat.

3. The heated air should move away from the area selected; that is, stagnant air spaces or immediate recirculation of the heated discharge air should be avoided.

4. Fan-cooled and ventilated rectifier equipment should be so oriented that the intake air comes from the cooler direction and the discharge is toward the warmer region as well as in line with the direction of exhaust from the space.

5. Natural baffling should be used to the best advantage, and consideration should be given to arrangement and orientation of all equipment in the space to get the best circulation of air.

6. The placing of metallic rectifiers in confined spaces should be avoided. Hot air may be trapped in such spaces and tends to accumulate at the top of a space. If such a location cannot be avoided, and other things are equal, the rectifiers should be placed low in the space, rather than high, but they should be above any bilge water splash level.

7. Equipment containing metallic rectifiers should not be placed adjacent to or in the immediate vicinity of other equipment that operates at temperatures appreciably higher than ambient temperature, nor should it be placed near any surface that is likely to reach high temperature. This precaution will prevent heat transfer to the equipment by radiation and conduction and will avoid reduction of the means to dissipate some of its own heat in the same manner.

8. Rectifier cells and stacks should not be painted or varnished.

9. Care should be taken to make sure the ambient temperature to which the equipment was designed is not exceeded. The instruction book should give the maximum ambient temperatures in which the rectifiers are to be placed.

10. Maintenance procedures should include an examination of equipment air openings and passages to make sure air flow has not been reduced by dust accumulation or by clogging of passages with foreign objects.

Regenerative d-c loads can have serious effects on rectifier power supplies unless the necessary precautions are taken. Loads which return power to the d-c system during some part of their operating cycle are known as regenerative. A rectifier will not absorb energy from the d-c bus. If there is nothing connected to the bus which will absorb power, a regenerative load will cause an increase in d-c bus voltage which could injure the rectifier stacks as well as cause a malfunctioning of the load which is regenerating.

Regeneration is of no concern if other nonregenerative loads capable of absorbing the regenerated power are always connected to the d-c bus. Fans and blowers are common loads of this type.

If other loads cannot be relied upon to absorb regenerated power, protective equipment must be provided to absorb it. In cases where regeneration is very frequent, and the power involved is small compared with the rectifier rating, a fixed resistor connected permanently to the d-c bus may be the most economical solution. In most cases, the power regenerated is sufficiently large to make it desirable to provide special switching equipment to connect an absorbing resistor to the bus only when regeneration occurs. Dynamic braking protective equipment may be used for this purpose.

Future Trends

The increasing use of magnetic amplifiers is turning the spotlight on germanium and silicon. Research and engineering are continually rolling back the limitations on these two new materials. Increasing acceptance by industry will inevitably reduce the cost until germanium and silicon may well dominate the metallic-rectifier field. However, the greater ruggedness and somewhat simpler application problems will enable selenium to retain its present leadership in many fields where maximum efficiency and high-temperature operation are not mandatory. For shipboard applications, the shock and vibration resistance and lack of maintenance requirements may well result in the practical monopoly by metallic rectifiers of all a-c to d-c power conversion requirements.

Time-Varying Analysis of a Guidance System

B. FRIEDLAND
NONMEMBER AIEE

IN A RECENT PAPER[1] the author presented a method of analyzing time-varying sampled-data systems. The essential features of this method are the characterization of components by triangular transmission matrices, and the use of ordinary matrix algebra to evaluate system performance. The purpose of the present paper is to illustrate by a typical problem how this method can be used in practical situations. The problem selected is that of steering a missile toward a target (both of which travel in a common plane) by means of a closed-loop control system. In addition to the intrinsic importance of this problem, the procedure one uses in the matrix analysis of the system is characteristic of a variety of problems. Since the objective here is illustration of a general method rather than detailed investigation of an actual system, simplifying assumptions will be used where convenient. In practice, these assumptions may have to be modified to obtain more precise results.

Nomenclature

θ_M = missile bearing
θ_T = target bearing
V_M = missile speed = constant
V_T = target speed = constant
$U = V_M - V_T$ = speed of closure
T = sampling interval
R_0 = initial range
$v_M = V_M T/R_0$ = normalized missile speed
$v_T = V_T T/R_0$ = normalized target speed
$u = UT/R_0$ = normalized speed of closure
δ = steering command
δ_s = sampled steering command
ϵ = system error

Description of System

The geometry of the steering problem is shown in Fig. 1. The quantities used therein and subsequently in this paper are defined in the nomenclature.

The kinematic differential equation for the missile bearing, θ_M, may be shown to be

$$r(t)\left[\frac{d\theta_M}{dt}+\frac{d\epsilon}{dt}\right]=V_M \sin\epsilon - V_T \sin(\theta_M+\epsilon-\theta_T) \quad (1)$$

This equation does not completely describe the missile kinematics, since the range $r(t)$ is given by another differential equation. By assuming that the actual trajectory of the missile does not depart from an ideal trajectory enough to affect the range, however, $r(t)$ may be approximated by an independent function of time, and the second equation is not needed.

The navigation plan used for the missile determines the error ϵ, which must be reduced to zero by the control system. A pure pursuit plan[2] is considered in the present paper (although the method may also be used for constant-bearing and proportional navigation). Ideally, in pure pursuit the missile should always be headed along the line of sight. Consequently, the system error is equal to the angle between the line of sight and the actual heading, i.e., $\epsilon = \epsilon$.

It is assumed that the control system has a device which obtains accurate samples of this error at a uniform sampling rate $1/T$. These samples are to shaped by a steering controller D, and passed into a zero-order hold (clamp) circuit. The output of this hold circuit is the steering command δ of the missile. The bearing of the missile is presumably related, through the dynamics of the steering engine and the missile, to the held steering command, by a linear differential equation.

Moreover, it is assumed that the system error is small, in which case, $\sin\epsilon \doteq \epsilon$, and that the missile and target are heading in approximately the same direction, in which case $\sin(\theta_M+\epsilon-\theta_T) \doteq \theta_M+\epsilon-\theta_T$. (The latter approximation is valid for a fixed target, or for the final portion of the pursuit trajectory for a moving target.) These approximations reduce equation 1 to the linearized equation

$$r(t)\frac{d\epsilon}{dt}-(V_M-V_T)\epsilon=-r(t)\frac{d\theta_M}{dt}-V_T\theta_M+V_T\theta_T \quad (2)$$

This equation and the previous considerations lead to the block diagram of Fig. 2 to represent the system.

To obtain the transfer function of the missile from δ to θ_M it is assumed that the rudder responds without delay to the steering command and hence that the rudder angle is identical to the steering angle δ; moreover it is assumed that if the rudder is set at a constant angle, the missile travels in a circle having a radius of curvature ρ inversely proportional to the rudder angle, i.e.,

$$\rho=\frac{K}{\delta}$$

Now, since

$$\rho\frac{d\theta_M}{dt}=V_M$$

$$\frac{d\theta_M}{dt}=\frac{V_M}{K}\delta$$

it is seen that the transfer function of the missile is

$$\frac{\theta_M(s)}{\Delta(s)}=\frac{V_M}{K}\frac{1}{s}$$

In a more realistic situation the rudder is incapable of responding instantaneously and the missile does not travel in exact circular arcs. The actual transfer function will thus be somewhat more complicated. The method one would use in this situation, however, is identical to that used for the simplified case, and only the latter is considered in order to make the calculations easier to follow.

Simplification of Block Diagram

If the initial error in launching the missile is small, and if the missile and target are traveling in approximately the same direction, then it is reasonable to assume that the actual trajectory will not differ greatly from a straight line, and hence that the range is adequately approximated by

$$r(t) \doteq R_0 - Ut \quad (3)$$

where R_0 is the initial range and $U = V_M - V_T$ = velocity of closure on target. With this assumption the minor loop in Fig. 2 is characterized by the differential equation

$$(R_0-Ut)\frac{d\epsilon}{dt}=x+U\epsilon \quad (4)$$

This is a first-order equation which may

Paper 58-82, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing December 11, 1957.

B. FRIEDLAND is with Columbia University, New York, N. Y.

The possibility of applying the matrix method for this investigation was suggested by Prof. John R. Ragazzini of the Columbia University Department of Electrical Engineering, to whom the author is grateful. This investigation was supported in part by the U. S. Air Force under Contract No. AF 18(600)-677, and monitored by the Office of Scientific Research, Air Research and Development Command.

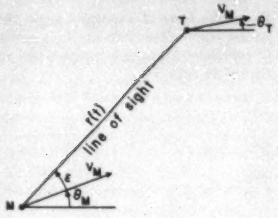


Fig. 1. Steering geometry

be solved by standard methods. The complete solution of the equation is

$$e(t) = \frac{1}{R_0 - Ut}\int x(t)\,dt + \frac{R_0}{R_0 - Ut}\,e_0$$

where e_0 is the initial launching error. Thus, if the initial error is treated as an input step of amplitude e_0, the minor loop may be replaced by an equivalent open loop. The system is thus represented by the block diagram of Fig. 3.

To apply the matrix method, the system must be represented by components all of which are separated by sampling switches. In general, this representation is accomplished by redrawing the block diagram and adding suitable fictiticus switches in series with switches already present. In the present system the target heading, θ_T, although actually continuous, may be represented as a sampled signal in the block diagram by following the sampled-signal by an appropriate holding circuit: for constant θ_T a zero-order hold is to be used, and for a target turning at a constant rate, a first-order hold is to be used. The block diagram of the system, which is now suitable for investigation by the matrix method is thus given by Fig. 4. Note that the missile bearing θ_M is an internal variable in the lower loop, and that the output of the redrawn system is the sampled steering command δ_s. Although θ_M will not be directly required in this computation, it may readily be obtained from the steering command by multiplying the transfer function of the plant, V_M/Ks, by the Laplace transform of the staircase-like steering function δ. The missile bearing at sampling instants is more readily computed by multiplying the pulse-transfer function of the plant, $(V_M T/K)(z^{-1}/1-z^{-1})$ by the Z-transform of the sampled steering command $\Delta_s(z)$.

Transmission Matrices of Components

The system has now been reduced to five components: D, F, G, H, and R^{-1}. The transmission matrices of these components must now be determined. Since G and H are fixed components, their transmission matrices are obtained directly from their pulse-transfer functions as explained in reference 1.

$$G(s) = \frac{V_T V_M}{K}\,\frac{1-e^{-sT}}{s}\,\frac{1}{s^2}$$

Thus, from a table of Z-transforms[2]

$$G(z) = \frac{V_T V_M T^2}{2K}\,\frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^3}$$

$$= \frac{v_M v_T R_0^2}{2K}\,\frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^3}$$

and the transmission matrix for G is

$$\mathbf{G} = \frac{v_M v_T R_0^2}{K}\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{2} & 0 & 0 & 0 & \cdots \\ \frac{3}{2} & \frac{1}{2} & 0 & 0 & \cdots \\ \frac{5}{2} & \frac{3}{2} & \frac{1}{2} & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}$$

For the case in which the target moves in a straight line, the hold in H is zero-order and thus

$$H(s) = V_T\,\frac{1-e^{-sT}}{s}\,\frac{1}{s}$$

from which

$$H(z) = V_T T\,\frac{z^{-1}}{1-z^{-1}} = v_T R_0\,\frac{z^{-1}}{1-z^{-1}}$$

and the transmission matrix for H is

$$\mathbf{H} = v_T R_0\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}$$

The component R^{-1} is a time-varying amplifier whose gain function is the reciprocal of the range, i.e.,

$$r^{-1}(nT) = \frac{1}{R_0 - nUT} = \frac{1}{R_0}\,\frac{1}{1-nu}$$

Hence the corresponding transmission matrix is

$$\mathbf{R}^{-1} = \frac{1}{R_0}\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & \frac{1}{1-u} & 0 & 0 & \cdots \\ 0 & 0 & \frac{1}{1-2u} & 0 & \cdots \\ 0 & 0 & 0 & \frac{1}{1-3u} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}$$

The elements of the transmission matrix of F are calculated from the general definition:[1] $f(n, k)$ is the amplitude of the output sample occurring at the nth sampling interval in response to single input sample applied at the kth sampling interval. The computation is illustrated graphically in Fig. 5. It is seen that an input sample of unit amplitude applied to the holding circuit produces a flat-topped pulse one sampling interval long as the output, shown in Fig. 5(B). Passing this pulse through the amplifier, whose gain is equal to the range results in a sloping topped pulse as shown in Fig. 5(C). This is finally integrated to give the waveform shown in Fig. 5(D). The amplitudes shown in the figure can readily be verified. It thus follows that

$$f(n, k) = 0 \quad \text{for } n < k+1$$

$$= \frac{v_M R_0^2}{K}\left[1 - \left(k + \frac{1}{2}\right)u\right] \text{for } n \geq k+1$$

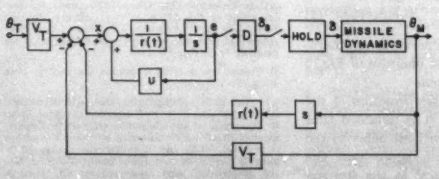


Fig. 2. Block diagram of system

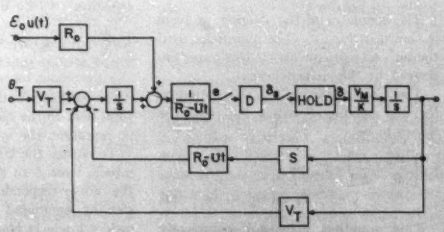


Fig. 3. Simplified block diagram

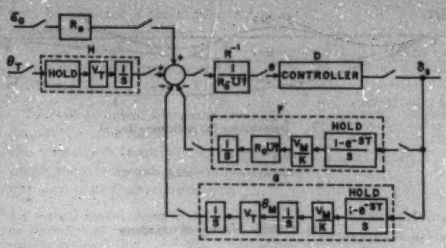


Fig. 4. Block diagram revised for matrix analysis

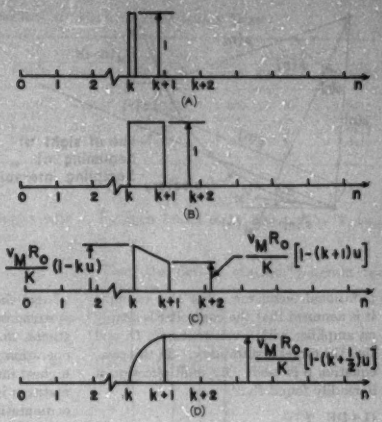


Fig. 5 (right). Computation of transmission matrix of F

A—Input to hold
B—Output from hold
C—Input to integrator
D—Output from integrator

and that the transmission matrix for F is

$$\mathbf{F}=\frac{v_M R_0}{K}\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1-\frac{1}{2}u & 0 & 0 & 0 & \cdots \\ 1-\frac{1}{2}u & 1-\frac{3}{2}u & 0 & 0 & \cdots \\ 1-\frac{1}{2}u & 1-\frac{3}{2}u & 1-\frac{5}{2}u & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Performance of System

From Fig. 4 the following relation between the (sampled) output steering command angle and the inputs is obtained:

$$\boldsymbol{\delta}_s=\mathbf{D}\mathbf{R}^{-1}(R_0\mathbf{e}_0+\mathbf{H}\boldsymbol{\theta}_T-\mathbf{F}\boldsymbol{\delta}_s-\mathbf{G}\boldsymbol{\delta}_s)$$

where R_0 is a constant, $\mathbf{D}$ is the transmission matrix of the steering controller, $\mathbf{R}^{-1}$, $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, are the transmission matrices defined in the previous section, and $\boldsymbol{\delta}_s$ and $\boldsymbol{\theta}_T$ are column vectors whose elements are the values of the corresponding variables at same intervals, and $\mathbf{e}_0=\text{col}[e_0,\ e_0,\ e_0,\ e_0,\ \ldots]$, e_0 being the initial launching error. Transposing the $\boldsymbol{\delta}_s$ terms to the left hand side of the equation and then premultiplying both sides by $(1+\mathbf{F}+\mathbf{G})^{-1}$ gives

$$\boldsymbol{\delta}_s=[1+\mathbf{D}\mathbf{R}^{-1}(\mathbf{F}+\mathbf{G})]^{-1}\mathbf{D}\mathbf{R}^{-1}(R_0\mathbf{e}_0+\mathbf{H}\boldsymbol{\theta}_T) \tag{5}$$

as the basic control equation.

Special Cases

1. Stationary Target. Since the transmission through G and H is proportional to the target velocity, the corresponding transmission matrices reduce to zero if the target is fixed. In this case equation 5 reduces to

$$\boldsymbol{\delta}_s=(1+\mathbf{D}\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{D}\mathbf{R}^{-1}R_0\mathbf{e}_0 \tag{6A}$$

2. Target Moving Along Initial Line of Sight. In this case it is convenient to chose as reference the initial missile direction and to measure angles from this line. Using this reference, if the target moves along the original line of sight then $\theta_T=e_0=$launching error and equation 5 becomes

$$\boldsymbol{\delta}_s=[1+\mathbf{D}\mathbf{R}^{-1}(\mathbf{F}+\mathbf{G})]^{-1}\mathbf{D}\mathbf{R}^{-1}(R_0+\mathbf{H})\mathbf{e}_0 \tag{6B}$$

Plotting Trajectories

As a result of assuming small launching errors, the rudder angles will also be small, and the radii of curvature of the circular arcs which constitute the trajectories will consequently be very large. To facilitate plotting the trajectories it is thus helpful to locate the position of the missile at the end of each sampling interval with respect to the line of sight at the beginning of that sampling interval. The appropriate relations are obtained with the aid of Fig. 6.

At the beginning of the nth sampling interval the missile is at point 1. At the end of the interval it has moved along a circular arc with radius $\rho(n)$ to point 2, and its distance from the line of sight is $d(n)$. It has moved along the line of sight a projected distance $l(n)$. It is shown in the appendix that for small errors $e(n)$, these are given by

$$d(n)=\frac{v_M R_0}{2}[A\delta_s(n)-2e(n)] \tag{7}$$

and

$$l(n)=v_M R_0\left[1-\frac{1}{2}e^2(n)+\frac{1}{2}Ae(n)\delta_s(n)-\frac{1}{6}A^2\delta_s^2(n)\right] \tag{8}$$

where

$$A=\frac{v_M R_0}{K}$$

If there is no error at the beginning of the sampling interval then the missile travels a distance of $v_M R_0$ on the line of sight in the interval. Thus the "foreshortening" due to the actual circular arc is

$$r(n)=v_M R_0-l(n)=v_M R_0\times\left[\frac{1}{2}e^2(n)-\frac{1}{2}Ae(n)\delta_s(n)+\frac{1}{6}A^2\delta_s^2(n)\right] \tag{9}$$

Since $\delta_s(n)$ is of the order of magnitude of $e(n)$, the foreshortening is roughly proportional to the square of the error, and is very small for small errors.

Trajectories of Uncompensated System with Stationary Target

Using equations 7 and 8, the trajectory of the missile may be plotted from a knowledge of the error signal e, and the (sampled) steering command δ_s. These computations are readily performed by substituting the transmission matrices derived in the previous sections into equation 6(A). The operations entail no more effort than the multiplication of matrices. Since the transmission matrices are all triangular, even the operation of inversion may be handled as easily as multiplication.[4] A simple example will serve to illustrate the general com-

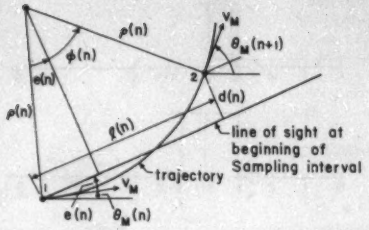


Fig. 6 (left). Detail of trajectory

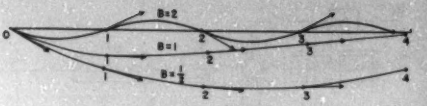


Fig. 7. Trajectories for stationary target

$V_M = 1/4$

$e_0 = 0.5$ radian $= 28.6$ degrees

putational technique. For this example it is assumed that the controller is simply an amplifier with a constant gain D, and that the target is stationary. In this case, equation 6 is used. By multiplication it is readily found that

$$[1+\mathbf{DR}^{-1}\mathbf{F}] = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ B\dfrac{1-u/2}{1-u} & 1 & 0 & 0 & \dots \\ B\dfrac{1-u/2}{1-2u} & B\dfrac{1-3u/2}{1-2u} & 1 & 0 & \dots \\ B\dfrac{1-u/2}{1-3u} & B\dfrac{1-3u/2}{1-3u} & B\dfrac{1-5u/2}{1-3u} & 0 & \dots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}$$

where $B = Dv_M R_0/K = AD =$ loop gain. Note that the nth row of the matrix is infinite when $1/u = n$, as a result of an apparently infinite gain in the forward loop at this sampling instant. The gain, R^{-1}, is inversely proportional to range, and is expected to be infinite at the $(1/u)$th sampling instant, since the range is ideally zero at this instant. Unless the target is actually intercepted at this instant, the range of course is not zero and the gain is not infinite. The effect of the approximation used here is most noticeable at the sampling instant the missile is expected to collide with the target. In general, the approximation will lead to an indication of infinite angular error. This indication should be interpreted as an actual angular error which is not small.

The matrix method leads to an approximation of the trajectory of the missile from the launching instant to the instant of expected interception which may be regarded as the termination time. If closed-loop control is preserved in the system after this time, a nonlinear problem results. On the other hand, if control is removed, then the missile simply continues along the circular arc it has been following. In either case the matrix method is no longer useful. Hence the system should be regarded as a terminal system, operating for $1/u$ sampling instants, and then changing its mode of operation. Only the first $1/u$ (or the next highest integer) rows of the transmission matrices involved need be carried in the computations.

The example will be continued for the case in which $u = 1/4$, i.e., four sampling intervals between launching and termination. While this is probably a smaller number of sampling intervals than might be used in practice, it serves to illustrate the method without excessively large matrices. With $u = 1/4$

$$\mathbf{1+DR}^{-1}\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 7B/6 & 1 & 0 & 0 & 0 \\ 7B/4 & 5B/4 & 1 & 0 & 0 \\ 7B/2 & 5B/2 & 3B/2 & 1 & 0 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

And the inverse is

$$(\mathbf{1+DR}^{-1}\mathbf{F})^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1.167B & 1 & 0 & 0 & 0 \\ -1.75B+1.458B^2 & -1.25B & 1 & 0 & 0 \\ -3.5B+5.54B^2-2.18B^3 & -2.5B+1.875B^2 & -1.5B & 1 & 0 \\ \infty & \infty & \infty & \infty & 1 \end{bmatrix}$$

Also with $u = 1/4$

$$\mathbf{R}^{-1}R_0 = \text{diag}\,[1,\ 1.333,\ 2,\ 4,\ \infty]$$

Hence

$$\mathbf{DR}^{-1}R_0\mathbf{e}_0 = \text{col}\,[1,\ 1.333,\ 2,\ 4,\ \infty]De_0$$

and by equation 6 $\boldsymbol{\delta}_z$ is obtained by premultiplying this column matrix by $(\mathbf{1+DR}^{-1}\mathbf{F})^{-1}$, i.e.,

$$\boldsymbol{\delta}_z = De_0 \begin{bmatrix} 1 \\ 1.333-1.167B \\ 2.000-3.417B+1.458B^2 \\ 4.00-9.83B+8.04B^2-2.18B^3 \\ \infty \end{bmatrix}$$

To obtain the trajectories note that since the controller is a constant gain amplifier,

$$e(n) = \frac{1}{D}\,\delta_z(n)$$

and equation 7 becomes

$$d(n) = \frac{v_M R_0}{2}[B-2]e(n)$$

and equation 9 becomes

$$f(n) = v_M R_0\left[\frac{1}{2} - \frac{1}{2}B + \frac{1}{6}B^2\right]e^2(n)$$

The numerical values of these quantities for three values of loop gain B, are given in Table I. The trajectories for an initial error of 0.5 radian are plotted in Fig. 7. If the range along the trajectories is compared with the approximation $r(t) = R_0 - Ut$, it will be noted that an error not exceeding 20% is made at the last sampling interval, and is substantially smaller at the earlier sampling intervals. However, an initial launching error of 0.5 radian $= 28.6$ degrees, chosen here to give a good picture, is not really small, and probably much larger than would be encountered in practice.

It is noted that with a loop gain $B = 1$, and only four sampling instants, the performance of the missile is quite good, and would probably result in a collision with a target of finite size. With a loop gain of 1/2 the system is too sluggish, and with a loop gain of 2 the missile "oscillates" about the initial line of sight, and is headed away from the target at the terminal instant; in this case the system appears to be too sensitive. From these observations one can conclude that the uncompensated system, for a reasonably large stationary target, may be expected to give satisfactory performance with the loop gain adjusted between 1.0 and about 1.5 in only four sampling instants. Naturally, superior performance

may be expected if the number of sampling intervals is increased.

Trajectories of Uncompensated System with Moving Target

In most practical situations the target cannot be considered stationary, and the trajectories resulting from a moving target must be investigated. At present two such cases shall be considered:

1. Target moves along initial line of sight in direction of missile: Target velocity = missile velocity.

2. Target moves along initial line of sight away from missile: Target velocity = 1/2 missile velocity.

These cases of course are not the only ones arising in practice, but they give an indication of how the missile might be expected to perform.

In both cases the missile velocity, initial range, and sampling interval are the same as used in the stationary target case, i.e., if the target were stationary the missile would reach the target in four sampling instants. Thus for case 1 the missile would reach the target in only two sampling instants, whereas in case 2 the missile would require eight sampling instants to reach the target. For both cases the loop gain B will be taken as unity.

The sequence of steering angles is obtained by means of equation 6(B). The computations are straightforward; hence only the results are given below:

Case 1. $v_M = 1/4$, $u = 1/2$, $v_T = -1/4$

$$\mathbf{e} = \text{col}[1.000,\ 0.250,\ \infty]\,e_0$$

$$\mathbf{d} = \text{col}[-0.125,\ -0.031,\ -]\,e_0 R_0$$

$$\mathbf{f} = \text{col}[0.041,\ 0.003,\ -]\,e_0^2 R_0$$

Case 2. $v_M = 1/4$, $u = 1/8$, $v_T = 1/8$

$$\mathbf{e} = \text{col}[1.000\ \ 0.143\ \ 0.000\ \ -0.029\ \ -0.036\ \ -0.038\ \ -0.039\ \ -0.040\ \ \infty]\,e_0$$

$$\mathbf{d} = \text{col}[-0.125\ \ -0.017\ \ 0.000\ \ +0.004\ \ 0.005\ \ 0.005\ \ 0.005\ \ -]\,e_0 R_0$$

$$\mathbf{f} = \text{col}[0.041\ \ \sim 0\ \ \sim 0\ \ \sim 0\ \ \sim 0\ \ \sim 0\ \ \sim 0\ \ \sim 0\ \ -]\,e_0^2 R_0$$

The corresponding trajectories are plotted in Fig. 8. It is seen that the missile is relatively insensitive to target motion, and that quality of performance appears to be governed primarily by the number of sampling intervals between launching and interception. Thus, when the target moves away from the missile, increasing the interception time, performance improves, while when the target moves toward the missile, decreasing the interception time, performance deteriorates. Similar results may be expected for targets moving at small angles with respect to the initial line of sight.

Table 1. Computations of Trajectories for Stationary Target

	B = 0.5			B = 1.0			B = 2.0		
n	$e(n)/e_0$	$d(n)/e_0R_0$	$f(n)/e_0^2R_0$	$e(n)/e_0$	$d(n)/e_0R_0$	$f(n)/e_0^2R_0$	$e(n)/e_0$	$d(n)/e_0^2R_0$	$f(n)/e_0^2R_0$
0	1.000	−0.168	0.0729	1.000	−0.125	0.0417	1.000	0	−0.0417
1	0.749	−0.140	0.0409	0.167	−0.021	0.0011	−1.000	0	+0.0417
2	0.656	−0.123	0.0314	0.041	−0.005	0.0000	+1.000	0	−0.0417
3	0.82	−0.15	0.049	0.03	−0.004	0.0000	−1.000	0	+0.0417
4	∞	—	—	∞	—	—	∞	—	—

Compensation of System

Although satisfactory performance is obtained when the steering controller, D, is simply an amplifier having a gain

$$D \triangleq \frac{K}{v_M R_0} = \frac{K}{V_M T}$$

there is always a finite error at the terminal instant. A system in which this error is intolerable may be compensated by use of an appropriate steering controller.

A suitable steering program for targets moving along initial line of sight is illustrated in Fig. 9. The steering angle is adjusted such that at the end of the second sampling interval the missile is located on, and heading along, the initial line of sight. With this steering program the missile will intercept any target moving along the line of sight, provided only that at least two sampling intervals are available and that the target does not have a speed advantage.

From geometrical considerations, if the launching error is small, then it is shown in the appendix that the steering commands must be

$$\delta(0) = \frac{3}{2}\frac{K}{v_M R_0}\,e(0)$$

$$\delta(1) = -\frac{1}{2}\frac{K}{v_M R_0}\,e(0)$$

$$\delta(2) = \delta(3) = \ldots = 0$$

Let the over-all transmission matrix between $\boldsymbol{\delta}$ and $\mathbf{e}_0$ be denoted by $\mathbf{T}$, i.e.,

$$\boldsymbol{\delta}_e = \mathbf{T}\mathbf{e}_0$$

Then the desired steering program requires that

$$\frac{K}{v_M R_0}\begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \mathbf{T}\begin{bmatrix} 1 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} e_0$$

This relation does not completely specify the elements of $\mathbf{T}$, of which there are $n(n+1)/2$ unknowns and only n equations. If the transmission between e_0 and δ_e is made time invariant, however (insuring that errors caused by disturbances later in the interception are also closed out) then the elements of $\mathbf{T}$ are uniquely determined. In particular, the equations for the elements of $\mathbf{T}$ are obtained from

$$\begin{bmatrix} t(0) & 0 & 0 & \ldots \\ t(1) & t(0) & 0 & \ldots \\ t(2) & t(1) & t(0) & \ldots \\ \cdot & \cdot & \cdot & \ldots \\ \cdot & \cdot & \cdot & \ldots \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \frac{K}{v_M R_0}\begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

i.e.,

$$t(0) = \frac{3}{2}\frac{K}{v_M R_0}$$

$$t(0) + t(1) = -\frac{1}{2}\frac{K}{v_M R_0}$$

$$t(0) + t(1) + t(2) = 0$$

$$\cdot$$

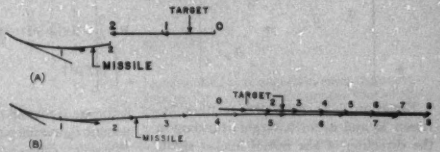
$$\cdot$$

$$\cdot$$

Fig. 8. Trajectories for moving target

A—Target moving toward missile, $V_T = -V_M$

B—Target moving away from missile $V_T = 1/2\ V_M$



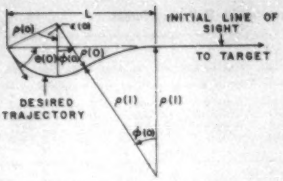


Fig. 9. Desired trajectory of compensated system

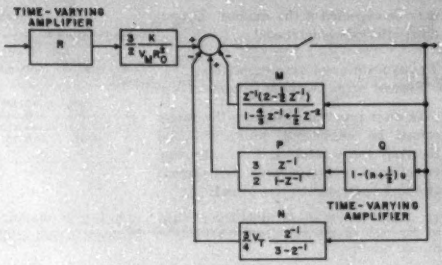


Fig. 10. Structure of compensator

Solving these equations gives

$$t(0) = \frac{3}{2}\frac{K}{v_M R_0}$$

$$t(1) = -2\frac{K}{v_M R_0}$$

$$t(2) = \frac{1}{2}\frac{K}{v_M R_0}$$

$$t(3) = t(4) = \ldots = 0$$

and

$$\mathbf{T} = \frac{K}{v_M R_0}\begin{bmatrix} \frac{3}{2} & 0 & 0 & 0 & \ldots \\ -2 & \frac{3}{2} & 0 & 0 & \ldots \\ \frac{1}{2} & -2 & \frac{3}{2} & 0 & \ldots \\ 0 & \frac{1}{2} & -2 & \frac{3}{2} & \ldots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}$$

Equation 6(B) gives the relation between $\bar{\delta}_t$ and $\mathbf{e}_0$ for a target traveling in the direction of the initial line of sight. From this equation

$$\mathbf{T} = [\mathbf{1} + \mathbf{D}\mathbf{R}^{-1}(\mathbf{F}+\mathbf{G})]^{-1}\mathbf{D}\mathbf{R}^{-1}(R_0\mathbf{+H})$$

and solving this equation for **D** gives

$$\mathbf{D} = [R_0\mathbf{T}^{-1} - \mathbf{F} + (\mathbf{H}\mathbf{T}^{-1} - \mathbf{G})]^{-1}\mathbf{R}$$

The components represented by $R_0\mathbf{T}^{-1}$ and $\mathbf{H}\mathbf{T}^{-1}-\mathbf{G}$ are fixed (time invariant) and may be characterized by pulse transfer functions as follows:

$$T(z) = \frac{K}{v_M R_0}\left(\frac{3}{2} - 2z^{-1} + \frac{1}{2}z^{-2}\right)$$

Thus

$$R_0 T^{-1}(z) = \frac{R_0}{T(z)} = \frac{v_M R_0^2}{K}\frac{1}{\frac{3}{2} - 2z^{-1} + \frac{1}{2}z^{-2}}$$

$$= \frac{2}{3}\frac{v_M R_0^2}{K}[1 + M(z)]$$

where

$$M(z) = \frac{z^{-1}\left(2 - \frac{1}{2}z^{-1}\right)}{1 - \frac{4}{3}z^{-1} + \frac{1}{2}z^{-2}}$$

and

$$H(z)T^{-1}(z) - G(z) = \frac{2}{3}\frac{v_M R_0^2}{K}N(z)$$

where

$$N(z) = \frac{3}{4}\frac{z^{-1}}{3 - z^{-1}}$$

Moreover, **F** may be written as

$$\mathbf{F} = \frac{2}{3}\frac{v_M R_0^2}{K}\mathbf{PQ}$$

where

$$\mathbf{P} = \frac{3}{2}\begin{bmatrix} 0 & 0 & 0 & 0 & \ldots \\ 1 & 0 & 0 & 0 & \ldots \\ 1 & 1 & 0 & 0 & \ldots \\ 1 & 1 & 1 & 0 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}$$

is the transmission matrix of a fixed component P, having a pulse transfer function $P(z) = 3z^{-1}/2(1-z^{-1})$ and

$$\mathbf{Q} = \text{diag}\left[1 - \frac{1}{2}u,\ 1 - \frac{3}{2}u,\ 1 - \frac{5}{2}u,\ 1 - \frac{7}{2}u,\ \ldots\right]$$

is the transmission matrix of an amplifier having a linearly varying gain, $1-(n+1/2)u$.

The transmission matrix of the compensator in terms of **M**, **N**, **P**, and **Q** may be written as

$$\mathbf{D} = \left[\frac{2}{3}\frac{v_M R_0^2}{K}(\mathbf{1+M-PQ+N})\right]^{-1}\mathbf{R}$$

$$= \frac{3}{2}\frac{K}{v_M R_0^2}(\mathbf{1+M-PQ+N})^{-1}\mathbf{R}$$

This relation may be obtained by the use of the relatively simple feedback structure shown in Fig. 10. Note that the slope of the gain of the amplifier in the second feedback path and the gain in the third loop, as well as the gain of the amplifier R, depend on the velocity of the target. Hence this quantity must be measured to insure perfect compensation. In addition, of course, the initial range R_0, and the missile velocity V_M must be known.

Conclusions

The computations indicate that the system may be expected to perform reasonably well with no compensation, if at least four sampling intervals are available during the interception period and if the loop gain is properly adjusted. Without compensation, however, an error will generally remain at the terminal instant. A compensator D, which insures that there will be no terminal error may be realized by using only fixed components, and amplifiers having time-varying gain.

Although the behavior of only a pure pursuit system has been investigated in this paper, the method may be used for constant bearing and proportional navigation systems, and for other systems of a similar nature. It may be summarized:

1. Obtain the differential equation relating the system error to the system output. (In general the equation will be nonlinear.)
2. Approximate this equation by a linear differential equation with variable coefficients by linearizing about an ideal trajectory.
3. Approximate the variable coefficients in the equation obtained in step 2 by assuming that the system follows the ideal trajectory.
4. Obtain the transfer function between the output of the system and the excitation.
5. Draw the block diagram of the system and manipulate it so that as many components as possible are separated by sampling switches.
6. Compute the transmission matrix of every component between a pair of sampling switches.
7. Derive the desired input-output relations in terms of these transmission matrices.

The foregoing steps having been completed, it is possible to compute the response of the system to any input. It is

also possible to solve for the transmission matrix of a compensator which will produce a required output (subject, of course, to physical realizability) for any given input.

Appendix

LOCATION OF MISSILE AT END OF SAMPLING INTERVAL; SEE FIG. 6

At the end of the nth sampling interval the missile is located at point 2. The displacement from the line of sight is given by

$$d(n) = \rho(n)\,[\cos\varepsilon(n) - \cos\varphi(n)] \qquad (10)$$

If $\varepsilon(n)$ and $\varphi(n)$ are small the following approximations may be used

$$\cos\varepsilon(n) \doteq 1 - \frac{1}{2}\varepsilon^2(n)$$

$$\cos\varphi(n) \doteq 1 - \frac{1}{2}\varphi^2(n)$$

and equation 10 becomes

$$d(n) = \frac{1}{2}\rho(n)\,[\varphi^2(n) - \varepsilon^2(n)]$$

$$= \frac{1}{2}\rho(n)\,[\varphi(n) + \varepsilon(n)]\,[\varphi(n) - \varepsilon(n)] \qquad (11)$$

Now it is seen that

$$\varphi(n) + \varepsilon(n) = \theta_M(n+1) - \theta_M(n) = \Delta\theta_M(n)$$

and

$$\rho(n) = \frac{K}{\delta_s(n)}$$

Hence equation 11 becomes

$$d(n) = \frac{K}{2\delta_s(n)}\Delta\theta_M(n)\,[\Delta\theta_M(n) - 2\varepsilon(n)] \qquad (12)$$

Now since the steering excitation consists of a staircase-like function which is integrated by the missile it can easily be shown that

$$\theta_M(n+1) = \frac{v_M R_0}{K}\sum_{k=0}^{n}\delta_s(n)$$

and that

$$\nabla\theta_M(n) = \frac{v_M R_0}{K}\,\delta_s(n) = A\delta_s(n)$$

Hence equation 12 becomes

$$d(n) = \frac{v_M R_0}{2}\,[A\delta_s(n) - 2\varepsilon(n)] \qquad (7)$$

The distance traveled by the missile along the line of sight is given by the following equation:

$$l(n) = \rho(n)\,[\sin\varepsilon(n) + \sin\varphi(n)] \qquad (13)$$

Assuming again that $\varepsilon(n)$ and $\varphi(n)$ are small the following approximations may be used:

$$\sin\varepsilon(n) \doteq \varepsilon(n) - \frac{\varepsilon^3(n)}{6}$$

$$\sin\varphi(n) \doteq \varphi(n) - \frac{\varphi^3(n)}{6}$$

Moreover, since the missile travels at constant speed,

$$\rho(n)\,[\varepsilon(n) + \varphi(n)] = V_M T$$

or

$$\rho(n) = \frac{v_M R_0}{\varepsilon(n) + \varphi(n)}$$

Hence equation 13 becomes

$$l(n) = v_M R_0\left\{1 - \frac{\varepsilon^3(n) + \varphi^3(n)}{6\,[\varepsilon(n) + \varphi(n)]}\right\}$$

$$= v_M R_0\left[1 - \frac{\varepsilon^3(n) - \varepsilon(n)\,\varphi(n) + \varphi^3(n)}{6}\right] \qquad (14)$$

Substituting

$$\varphi(n) = A\delta_s(n) - \varepsilon(n)$$

into equation 14 gives

$$l(n) = v_M R_0\left[1 - \frac{1}{2}\varepsilon^3(n) + \frac{1}{2}A\varepsilon(n)\delta_s(n) - \frac{1}{6}A^2\delta_s^2(n)\right]$$

which is equation 8 of the text.

IDEAL STEERING PROGRAM; SEE FIG. 9

Since the circular arc segments of the desired trajectory must be tangent, then the following is true:

$$[\rho(1) + \rho(0)]^2 = [\rho(1) + \rho(0)\cos\varepsilon(0)]^2 + [L - \rho(0)\sin\varepsilon(0)]^2$$

or, simplifying,

$$\rho(0) = \frac{L\,[L - 2\rho(0)\sin\varepsilon(0)]}{2\rho(0)\,[1 - \cos\varepsilon(0)]} \qquad (15)$$

also

$$\rho(0)\,[\sin\varepsilon(0) + \sin\varphi(0)] + \rho(1)\sin\varphi(0) = L \qquad (16)$$

The missile travels an equal length of time on each segment of the trajectory, thus

$$\rho(0)\,[\varepsilon(0) + \varphi(0)] = \rho(1)\,\varphi(0) \qquad (17)$$

For small angles $\varepsilon(0)$ and $\varphi(0)$ equations 15 and 16 become, respectively

$$\rho(0) \doteq \frac{L\,[L - 2\rho(0)\,\varepsilon(0)]}{\rho(0)\,\varepsilon^2(0)} \qquad (18)$$

and

$$\rho(0)\,[\varepsilon(0) + \varphi(0)] + \rho(1)\,\varphi(0) = L \qquad (19)$$

Simultaneous solution of these equations yields

$$\rho(0) = \frac{L}{3\varepsilon(0)}$$

$$\rho(1) = \frac{L}{\varepsilon(0)}$$

Now if the launching error is small then it is reasonable to assume that

$$L \doteq 2V_M T = 2v_M R_0$$

from which

$$\rho(0) \doteq \frac{2}{3}\,\frac{v_M R_0}{\varepsilon(0)}$$

$$\rho(1) \doteq 2\,\frac{v_M R_0}{\varepsilon(0)}$$

But $\delta_s(n) = K/\rho(n)$. Thus

$$\delta_s(0) = \frac{3}{2}\,\frac{K}{v_M R_0}\,\varepsilon(0)$$

$$\delta_s(1) = -\frac{1}{2}\,\frac{K}{v_M R_0}\,\varepsilon(0)$$

Note that $\delta_s(1)$ is negative, since the curvature of the second arc is opposite that of the first.

References

1. A TECHNIQUE FOR THE ANALYSIS OF TIME-VARYING SAMPLED-DATA SYSTEMS, B. Friedland. *AIEE Transactions*, vol. 75, pt. II, 1956 (Jan. 1957 section), pp. 407–14.

2. GUIDANCE (book), A. S. Locke, editor. "The Analysis of Flight Paths," D. Van Nostrand Company, Inc., Princeton, N. J., 1955, chap. 12.

3. ANALYSIS AND SYNTHESIS OF SAMPLED-DATA CONTROL SYSTEMS, E. I. Jury. *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

4. THEORY OF TIME-VARYING SAMPLED-DATA SYSTEMS, B. Friedland. *Technical Report T-19/B*, Electronics Research Laboratories, Columbia University, New York, N. Y., Apr. 1957.

An Analysis of a Servomechanism with Backlash by the Ritz-Galerkin Method

KATSUHIKO OGATA
ASSOCIATE MEMBER AIEE

C. P. ATKINSON
NONMEMBER AIEE

Synopsis: The Ritz-Galerkin method is used to develop the equations for finding frequency response curves for a servomechanism with backlash. Response curves are presented for free oscillations of the system for three values of gain. The Ritz-Galerkin method is used to find the hunting frequency and hunting amplitude of such a system. Favorable agreement was found between an experimental check with an electronic analog computer and the calculated results.

THIS PAPER presents the results of a theoretical investigation (with experimental confirmation by analog computer) of a nonlinear servomechanism with backlash. Equations for finding the frequency response curves of the backlash system are developed by use of the Ritz-Galerkin method.[1,2] In the absence of a driving force, such systems are known to exhibit hunting phenomena in which the system oscillates with a certain amplitude and frequency. An analytical method is developed for finding the hunting frequency and hunting amplitude of such a system.

Several methods have been developed for analyzing this system. Vallese[3,4] uses the method of Kryloff and Bogoliuboff to obtain hunting frequency and amplitude. Nichols[5] used the describing function method to find the hunting frequency and amplitude of a velocity lag servomechanism. Greif[6] obtained the describing function of a backlash or hysteresis element. Hsia[7] analyzed a servomechanism with backlash by a graphical method. Tustin[8] used Nyquist diagrams to analyze servomechanism with backlash. Satyendra[9] investigated the effect of backlash and coulomb friction on the stability of a feedback control system by the describing function method. Halstenberg used the describing function method to analyze control systems with hysteresis and nonlinear gain.

In this paper the Ritz-Galerkin method was used for finding the hunting frequency and hunting amplitude of a second order servomechanism with backlash. Good agreement was obtained between the theoretical results developed in the paper and an experimental check on an electronic analog computer.

Nomenclature

a = the half period of $m(t)$
$A = M \sin \phi$
A_n = amplitude of the sine term of the nth component of the assumed solution $m(t)$
$B = M \cos \phi$
B_n = amplitude of the cosine term of the nth component of the assumed solution $m(t)$
c, $c(t)$ = output of the system
e, $e(t)$ = error signal
F = a function of a variable, its derivatives and time
K = gain
m_0 = half width of the backlash zone; see Fig. 4
m, $m(t)$ = output of the motor
$\dot{m}$, $\ddot{m}$ = time derivatives of $m(t)$
M = amplitude of the assumed solution for $m(t)$
n = an integer
r, $r(t)$ = input to the system
R = amplitude of the assumed $r(t) = R \cos \omega t$
s = Laplace transform operator
t = time
T = time constant of the motor
ϕ = phase angle between $r(t)$ and $m(t)$
$\pi = 3.1416$
τ = time interval over which $c(t)$ is constant; see Fig. 4
ω = circular frequency of oscillation

The System Studied and the Method Used for Studying It

The System Studied

The system studied for this paper is shown schematically in Fig. 1, which shows a feedback control system consisting of an amplifier, a motor, and a gear train. The position of the output is fed back to the input to generate the error signal which will activate the system. Fig. 2 shows the block diagram for the system. The box containing the transfer function $K/s(Ts+1)$ represents the amplifier and the motor characteristics. The second box represents the backlash of the gear system. The inertia of the gears is considered negligible compared with that of the motor.

The Ritz-Galerkin Method

The Ritz-Galerkin method can be outlined briefly as follows: Given a differential equation

$$F(\ddot{m}, \dot{m}, m, t) = 0 \tag{1}$$

In the absence of an exact solution, the assumption is made that

$$m(t) = \sum_{n=1}^{n} (A_n \sin n\omega t + B_n \cos n\omega t)$$

The Ritz-Galerkin method then requires the calculation of the integrals of the form

$$\int_0^{2a} F(\ddot{m}, \dot{m}, m, t) \sin n\omega t \, dt = 0$$

$$\int_0^{2a} F(\ddot{m}, \dot{m}, m, t) \cos n\omega t \, dt = 0 \tag{2}$$

where $2a = 2\pi/\omega$, the longest period of the components of the solution $m(t)$.

The evaluation of the integrals in equation 2 yields the Ritz equations which are algebraic equations for the solution of the constants A_n and B_n in terms of ω. From these equations we can plot frequency response curves.

Application of the Ritz-Galerkin Method to the Nonlinear Servomechanism

The general equation describing the dynamics of the servo system of Fig. 2 is

$$T\ddot{m} + \dot{m} = Ke \tag{3}$$

where $m(t)$ is the output of the motor and where the error signal e is

$$e = r - c \tag{4}$$

where r is the input to the system

$$r = R \cos \omega t \tag{5}$$

and c is the output of the system. T and K are constants of the system. The assumed solution of equation 3 is

$$m(t) = M \cos(\omega t - \phi)$$
$$= A \sin \omega t + B \cos \omega t \tag{6}$$

where

$$A = M \sin \phi; \quad B = M \cos \phi$$

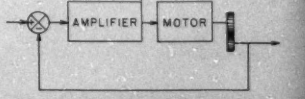


Fig. 1. Servomechanism consisting of amplifier, motor, and gear train

---

Paper 58-86, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted September 3, 1957; made available for printing November 19, 1957.

KATSUHIKO OGATA is with the University of Minnesota, Minneapolis, Minn.; and C. P. ATKINSON is with the University of California, Berkeley, Calif.

The research for this work was sponsored by the National Science Foundation.

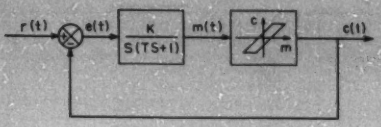


Fig. 2 (left). Block diagram of system shown in Fig. 1

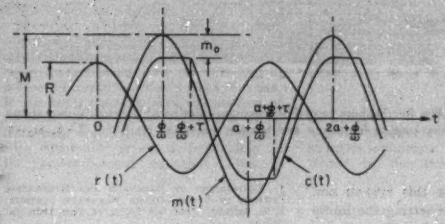


Fig. 3 (left). Phase relation between r(t) and m(t), and the relation between m(t) and c(t)

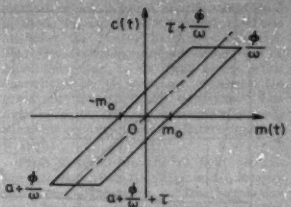


Fig. 4. Input-output characteristics of backlash nonlinearity

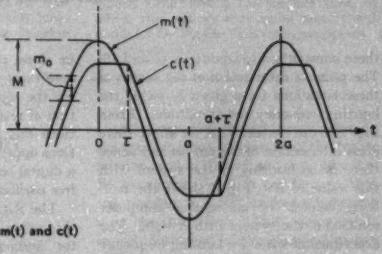


Fig. 5 (right). Relation between m(t) and c(t) for free oscillation

The relation between the output of the system $c(t)$, after the backlash element, and $m(t)$ the output of the motor, before the backlash element, is given as follows (see Figs. 3 and 4 for a graphic description of these limits):

$$c(t)=M-m_0 \text{ for } \frac{\phi}{\omega}\leqslant t\leqslant \tau+\frac{\phi}{\omega}$$

$$c(t)=m(t)+m_0 \text{ for } \frac{\phi}{\omega}+\tau\leqslant t\leqslant a+\frac{\phi}{\omega}$$

$$c(t)=-M+m_0 \text{ for } \frac{\phi}{\omega}+a\leqslant t\leqslant a+\tau+\frac{\phi}{\omega}$$

$$c(t)=m(t)-m_0 \text{ for } \frac{\phi}{\omega}+a+\tau\leqslant t\leqslant 2a+\frac{\phi}{\omega} \quad (7)$$

where m_0 is the half width of the backlash zone as shown in Fig. 4, M is the amplitude of $m(t)$ and τ is the time interval over which $c(t)$ is constant (see Fig. 4).

So the Ritz integrals (equation 2) for this system are

$$\int_{\frac{\phi}{\omega}}^{2a+\frac{\phi}{\omega}} [T\ddot{m}+\dot{m}-KR\cos\omega t+Kc(t)]\sin\omega t\,dt=0 \quad (8)$$

$$\int_{\frac{\phi}{\omega}}^{2a+\frac{\phi}{\omega}} [T\ddot{m}+\dot{m}-KR\cos\omega t+Kc(t)]\cos\omega t\,dt=0 \quad (9)$$

where $a=\pi/\omega$.

Equation 8 yields the first Ritz equation

$$-AT\omega\pi-B\pi+\frac{2KM}{\omega}[\cos\phi-\cos\times(\phi+\omega\tau)]+\frac{4Km_0}{\omega}\cos(\phi+\omega\tau)+KA\times(a-\tau)-\frac{KA}{2\omega}[\sin 2\phi-\sin 2(\phi+\omega\tau)]+\frac{KB}{\omega}[\sin^2\phi-\sin^2(\phi+\omega\tau)]=0 \quad (10)$$

Equation 9 yields the second Ritz equation

$$-BT\omega\pi+A\pi-KR\frac{\pi}{\omega}+\frac{KM}{\omega}[2\sin(\phi+\omega\tau)-2\sin\phi]-\frac{4Km_0}{\omega}\sin(\phi+\omega\tau)+\frac{KA}{\omega}[\sin^2\phi-\sin^2(\phi+\omega\tau)]+KB(a-\tau)+\frac{KB}{2\omega}[\sin 2\phi-\sin 2(\phi+\omega\tau)]=0 \quad (11)$$

At the hunting frequency the work done during one period will be zero. If this were not so the steady state oscillation could not be sustained. The hunting frequency is the frequency of the limit cycle. The work done at the hunting frequency over one period can be expressed as follows:

$$\int_0^{2a} F(\dot{m},m)\,dm=0 \quad (12)$$

Where $2a$ is the period, $F(\dot{m},m)$ is equivalent to a force and m is equivalent to a displacement.

From equations 3 and 4, equation 12 becomes

$$\int_0^{2a} [\dot{m}+Kc(t)]\sin\omega t\,dt=0 \quad (13)$$

where $c(t)$ is a function of m as shown in equation 7.

The final result of the integration of equation 13 is

$$\frac{\pi}{K}-\frac{4m_0}{\omega M}\left(1-\frac{m_0}{M}\right)=0 \quad (14)$$

If $m(t)$ is the solution at hunting then it should satisfy equation 14 and the Ritz equation 9 when $R=0$. If $m(t)$ is assumed to be (see Fig. 5 for a graphic description of this special case)

$$m(t)=M\cos\omega t \quad (15)$$

the Ritz equation 9 becomes

$$\left(1-\frac{T\omega^2}{K}\right)\tau=\omega\tau-\frac{1}{2}\sin 2\omega\tau \quad (16)$$

where

$$\omega\tau=\cos^{-1}\left(1-\frac{2m_0}{M}\right) \quad (17)$$

The simultaneous solution of equations 14 and 16 yields the hunting frequency and hunting amplitude.

Theoretical and Experimental Results

Equations 14 and 16 were solved for the following system constants: K, the gain of the system was set equal to the three values 1, 5, 10; T, the time constant of the motor was assumed equal to one; m_0, the half width of the backlash zone of Fig. 4 was set to one. Fig. 6 shows curves resulting from the substitution of

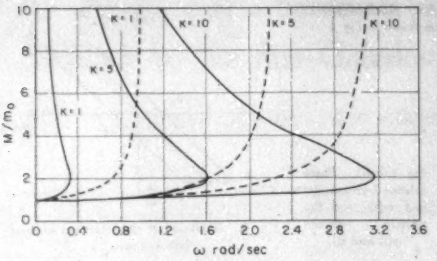


Fig. 6. Curves corresponding to equations 14 and 16

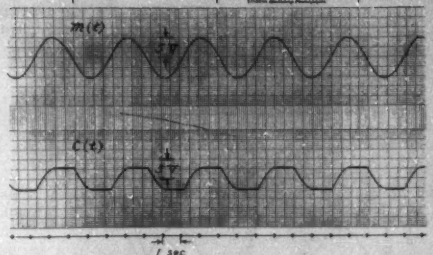


Fig. 8. Sanborn record of computer solution of backlash system showing hunting frequency and hunting amplitude for K=5, T=1, m_0=1

these constants into equations 14 and 16. The point of intersection of the curves of these equations for a given K yields the hunting frequency and amplitude. From Fig. 6 it can be seen that for $K=1$ there is no intersection of the curves and hence there is no hunting in the system with this value of K. Fig. 7 shows the Sanborn records of an analog computer solution of the system with $K=10$. The experimental value for hunting frequency is $\omega=2.69$ radians per second. The hunting amplitude M is equal to 3.46. Fig. 6 indicates the theoretical value for the hunting frequency as $\omega=2.75$ radians per second and the hunting amplitude of $M=3.35$. Fig. 8 shows the computer solution for $K=5$ and yields a hunting frequency $\omega=1.61$ radians per second. The hunting amplitude in Fig. 8 is $M=2.21$. Fig. 6 for $K=5$ yields $\omega=1.6$ and $M=2.0$.

Conclusions

The Ritz-Galerkin method may be used to develop equations for finding the fundamental response curves for a nonlinear servomechanism. The evaluation of the Ritz integrals over a period must be done piecewise for this system and care must be taken in setting the limits for the separate integrals. The calculation of response curves for forced oscillations was not attempted since the equations appear to be more easily handled by a digital computer. Response curves for free oscillations were calculated.

The Ritz-Galerkin method as used in this paper is another analytical method for finding the hunting frequency and hunting amplitude of a nonlinear servomechanism with backlash. The results from this method for finding hunting frequency and hunting amplitude compare favorably with experimental results gotten from an electronic analog computer.

References

1. Non-Linear Vibration Problems Treated by the Averaging Method of W. Ritz, K. Klotter. *Proceedings*, First U. S. National Congress of Applied Mechanics, ASME. J. W. Edwards, Ann Arbor, Mich., 1952, pp. 125–31.

2. Steady State Vibrations in Systems Having Arbitrary Restoring and Arbitrary Damping Forces, K. Klotter. *Proceedings*, Symposium on Non-Linear Circuit Analysis, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1953, pp. 234–58.

3. Analysis of Backlash in Feedback Control Systems With One Degree of Freedom, L. M. Vallese. *AIEE Transactions*, vol. 74, pt. II, Mar. 1955, pp. 1–4.

4. Comparison of Backlash and Hysteresis Effects in Second-Order Feedback Systems, L. M. Vallese. *Ibid.*, vol. 75, pt. II, Sept. 1956, pp. 240–43.

5. Backlash in a Velocity Lag Servomechanism, Nathaniel B. Nichols. *Ibid.*, vol. 72, pt. II, 1953 (Jan. 1954 section), pp. 402–67.

6. Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities, H. D. Greif. *Ibid.*, vol. 72, pt. II, Sept. 1953, pp. 243–48.

7. A Graphical Analysis for Non-Linear Systems, P. S. Hsia. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. II, 1952, pp. 125–34.

8. The Effects of Backlash and of Speed Dependent Friction on the Stability of Closed-Cycle Control Systems, A. Tustin. *Journal*, Institution of Electrical Engineers, vol. 94, pt. IIA, 1947, pp. 143–51.

9. Describing Functions Representing the Effect of Inertia, Backlash, and Coulomb Friction on the Stability of an Automatic Control System—I, K. N. Satyendra. *AIEE Transactions*, vol. 75, pt. II, Sept. 1956, pp. 243–49.

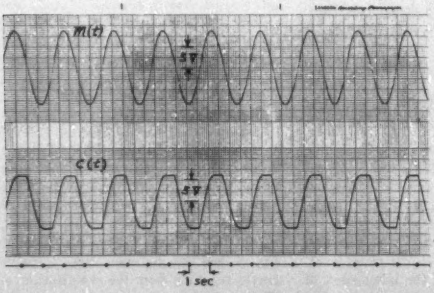


Fig. 7. Sanborn record of computer solution of backlash system showing hunting frequency and hunting amplitude for K=10, T=1, m_0=1

Discussion

R. L. Cosgriff (Ohio State University, Columbus, Ohio): The authors are to be congratulated for presenting this paper for it illustrates an old method of analysis not widely known today. However, there are several points in this development that should be amplified in order to correlate this method of analysis with others currently in use.

First the final results of the analysis are exactly equivalent to those published by Johnson;[1] for the results using the harmonic balance equations are always the same as those obtained by the describing function (nonlinear gain) method. The difference lies only in the method of achieving these results. Here the authors have kept the representation in the real domain while those using the nonlinear gain approach use the complex notation for representing the system variables.

The two Ritz (harmonic balance) equations correspond to the real and imaginary parts of the complex expression for the system response.[2,3]

In general, the complex representation is somewhat easier to use than the har-

monic balance approach in that the only problem involved is that associated with the nonlinear elements. Essentially, the integration problem is simplified. It would appear that the Ritz method for obtaining harmonics other than the first is easier than that used in references 2 and 3 of this discussion. However, it also appears that certain methods employing complex representation is easier than either of these two methods. In any case harmonic solutions are very tedious.

In order to solve for the limit cycle conditions (amplitude and frequency of oscillation) the authors used the two Ritz equations and then used the principle of balance of energy to obtain a third equation. If this latter equation were independent of the first two equations, a solution would have been accidental. However, the first Ritz equation[4,5] is exactly equivalent to the "power balance" equation. Thus, in effect, the authors are using only the two Ritz equations.

It is interesting to note that the second Ritz equation is equivalent to a second principle,[4,5] namely "balance of reactive power."

References

1. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements. **Calvin Johnson.** *AIEE Transactions*, vol. 71, pt. II, pp. 169–81.

2. Nonlinear Mechanics (book). **N. Minorsky.** J. W. Edwards, Ann Arbor, Mich., 1947, chap. XII.

3. Introduction to Nonlinear Mechanics (book). **N. Kryloff, N. Bogoliaboff.** Kiev, USSR, 1937, chaps. X, XI, XII.

K. Ogata and C. P. Atkinson: The authors wish to thank Prof. Cosgriff for his constructive comments. As the authors mentioned in the text, the Ritz-Galerkin method, which is based on the variational method, is one of the several methods available for finding harmonic solutions. The Ritz-Galerkin method seems to be more convenient to use than others when characteristics of nonlinear elements are given in the form of equations. The authors agree that, in general, harmonic solutions are very tedious whatever method is used.

Prof. Cosgriff brought to light an important fact that the Ritz equations are equivalent to "power balance" equation and "reactive power balance" equation. The authors used equations 13 and 9 for finding the hunting frequencies and amplitudes of the backlash systems. However, it is true that equations 8 and 13 are equivalent, as was mentioned by the discusser. Therefore it can be said that, in effect, the authors were using equations 8 and 9 for finding the limit cycle conditions.

Transient Protection Conditions in Pipe-Line Stations

E. B. TURNER
ASSOCIATE MEMBER AIEE

PROTECTION of electric equipment transient disturbances caused by lightning or power-line switching is a subject which is of concern in every industry which was electric equipment. To provide a more complete picture of the considerations involved, the general situation is examined as well as the specialized conditions existing in pipe-line installations and other petroleum industry applications.

Lightning and Surge Protection

In general, protection must be provided in two classifications:

1. Protection from direct strokes of lightning.
2. Protection from traveling waves.

Protection against direct strokes is not covered in this paper, since most installations with which the author is concerned have relatively limited exposure to direct stroke. The area which is exposed is usually the transmission system, and this is protected by the utility company. Therefore, protection against traveling waves is the problem of greater concern in this paper.

These traveling waves can be the result of direct lightning strokes out on the line, insulator flashover, or induced by lightning stroke on adjacent lines. The fourth possibility is switching surges which cause a steepfront wave to be transmitted along the line. When this traveling wave reaches a point in the system which has a different surge impedance, reflections result which have magnitudes of the order described by the following equation (see Appendix I)

$$E_t = 2E_i \frac{Z_t}{Z_i + Z_t}$$

where

E_i = incoming voltage
E_t = outgoing (refracted) voltage
Z_i = surge impedance of incoming line
Z_t = surge impedance of outgoing line

In almost every instance, oil pipe-line stations have station lightning-arrester protection. The station is usually supplied by step-down transformers reducing the incoming line voltage from the transmission level to either 2,400 or 4,160 volts. By examining Fig. 1, it can be seen that the exposure to direct strokes of lightning is at the substation. This equipment is usually furnished by the electrical utility with the proper lightning protection.

When a traveling wave is imposed upon the high side of a transformer, different magnitudes of transient voltages are encountered, depending upon the size and connections of the transformer. The resulting voltage which appears at the motor terminals and is transmitted through a given transformer depends upon the kva size of the transformer and voltage level of the motor. It appears to be independent of the transformer ratio if the ratio exceeds four to one.[1] Fig. 2 illustrates the maximum expected magnitudes of surge voltage appearing at the machine terminals. It should be noted that maximum magnitudes are dependent upon the use of lightning arresters and upon the high-side connections.[1]

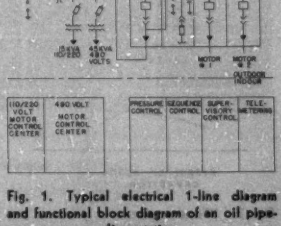


Fig. 1. Typical electrical 1-line diagram and functional block diagram of an oil pipe-line station

Paper 57-990, recommended by the AIEE Petroleum Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Conference of the Petroleum Industry, Philadelphia, Pa., September 9–11, 1957. Manuscript submitted June 26, 1957; made available for printing October 9, 1957.

E. B. Turner is with General Electric Company, Schenectady, N. Y.

The author wishes to thank P. D. Woodruff for his assistance in conducting and analyzing the oscillographic tests, and also the Plantation Pipe Line Company for making available personnel and facilities to assist in conducting the oscillographic tests on a pump motor at their Doraville station.

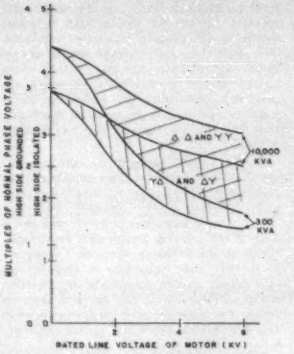


Fig. 2. Expected magnitudes of surge voltages transmitted through transformers

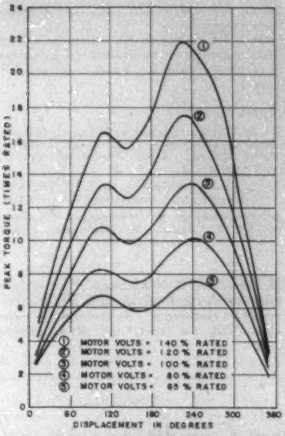

PEAK TORQUE (TIMES RATED)
DISPLACEMENT IN DEGREES
① MOTOR VOLTS = 140 % RATED
② MOTOR VOLTS = 120 % RATED
③ MOTOR VOLTS = 100 % RATED
④ MOTOR VOLTS = 80 % RATED
⑤ MOTOR VOLTS = 65 % RATED

Fig. 4. Expected peak torque developed in motor versus angle of displacement from power system during reclosure, for various motor voltages

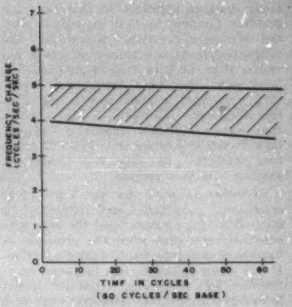


Fig. 6. Rate of frequency change

Protection of Machines: Traveling Waves

Protection of machines may be thought of as consisting of two entirely separate problems:

1. Limiting the maximum voltage to ground imposed upon the motor windings.
2. Limiting the turn-to-turn voltage in the motor windings; see Appendix II.

The wave approaching the motor has a maximum voltage which is reduced to an acceptable level by the use of lightning arresters connected line to ground.

The insulation of a motor between turns of the same coil is subjected to relatively small voltage differentials in normal service. The steepness of a traveling wave can instantaneously cause excessive differential voltages to exist, the magnitude of this voltage depends upon the rate of rise of voltage and the length of the motor coils. Proper protection can be provided by using a capacitor connected line to ground in parallel with the arrester. The capacitance reduces the steepness of the wavefront and thus limits the turn-to-turn voltage to a safer value; Fig. 3.

Therefore, the proper protection for a machine connected to a power system through a transformer is a surge capacitance in parallel with a lightning arrester.

The ideal location for surge capacitors and lightning arresters is at the terminals of the machine being protected. However, due to the hazardous atmosphere usually found at the machine terminals, it is not practical in the petroleum industry applications to locate this equipment at that point. It should not be concluded, therefore, that this protection should be eliminated, since it is an essential part of the electrical protection system. The proper place under these circumstances is to locate the surge capacitor and lightning arrester on the station bus. The length of the cable connection from the bus to the lightning arrester and surge capacitor should be kept to a minimum.

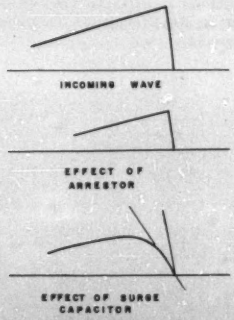


Fig. 3. Effects produced by transient protective equipment

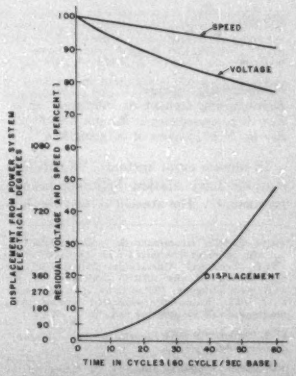


Fig. 5. Test results of power interruption on 1,250-hp motor operating with 180 kva of power-factor corrective capacitors

Motor Transients: Torque and Current

Power interruptions of approximately 20-cycle duration can cause undesirable transient conditions in operating motors. Interruptions of this duration can be caused by instantaneous reclosing of circuit breakers on the transmission system serving the station. These undesirable transients are excessively high currents and torques which result from reapplying power out of phase with the residually generated voltage of the operating motors (Appendix III). Factory tests have shown that magnitudes of 12 times rated line current and 20 times rated torque can be approached; see Fig. 4.

Oscillographic tests were conducted to determine the actual conditions which exist when the power is removed from an operating motor. These tests showed

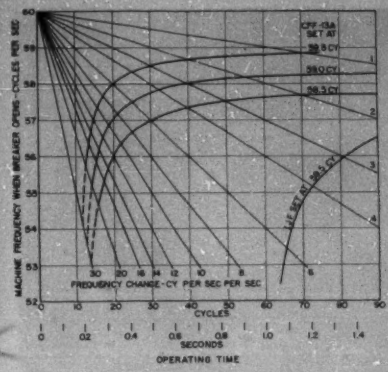
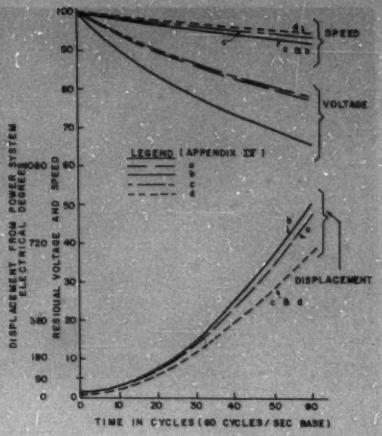


Fig. 7. Under frequency relay characteristics including total operating time

Fig. 9 (right). Test results of power interruption on 1,250-hp motor
A—Power - factor corrective capacitors connected, discharge valve open
B—Power - factor corrective capacitors disconnected, discharge valve open
C—Power - factor corrective capacitors connected, discharge valve closed
D—Power - factor corrective capacitors disconnected, discharge valve closed



that the residually generated voltage decayed very slowly. The rate of change of frequency was also very low; see Appendix IV.

The oscillographic tests were conducted on an actual pipe-line installation, with the use of a totally enclosed fan-cooled explosionproof motor rated 1,250 hp (horsepower), 3,600 rpm, 2,300 volts. The motor was operating at rated speed when power was removed from the motor, and measurements were taken of the residually generated voltage and frequency, speed of the motor, and the power system frequency, for use as a reference. For the condition when the motor and pump are operating normally with power-factor corrective capacitors (Fig. 5), after 20 cycles of power interruption the voltage is 90% of rated, the speed is 97%, and the residual voltage has shifted behind the power system by approximately 118 electrical degrees. This is an undesirable situation, and power should not be reapplied under these conditions.

Although most pipe-line stations have undervoltage protection with time delays of 1/2 to 1 second, even instantaneous undervoltage could not operate to disconnect the motor due to the high residual voltage maintained by the motors. In this case also, the rate of frequency change (Fig. 6) is so low that operation of the breaker by underfrequency relaying is also excluded (Fig. 7).

The oscillographic tests (Figs. 10(A) through D) did not factor in other loads connected to the power system supplying the motors. These system effects

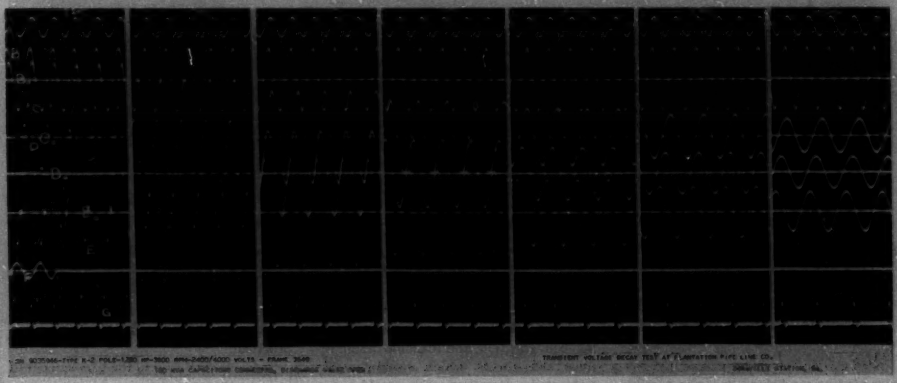
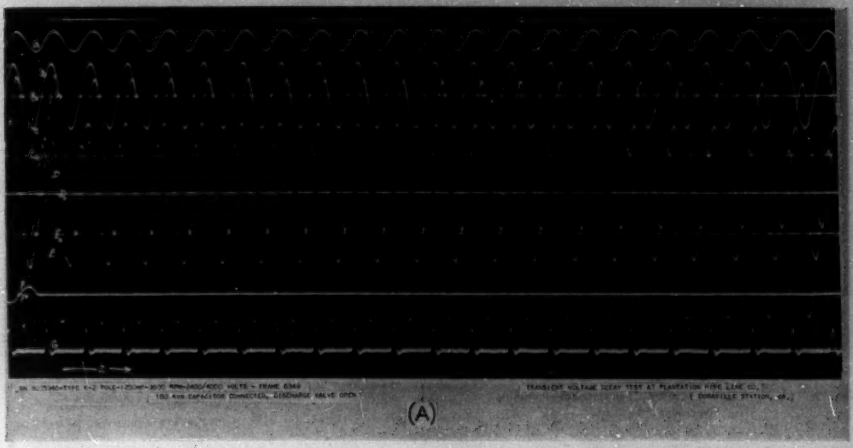
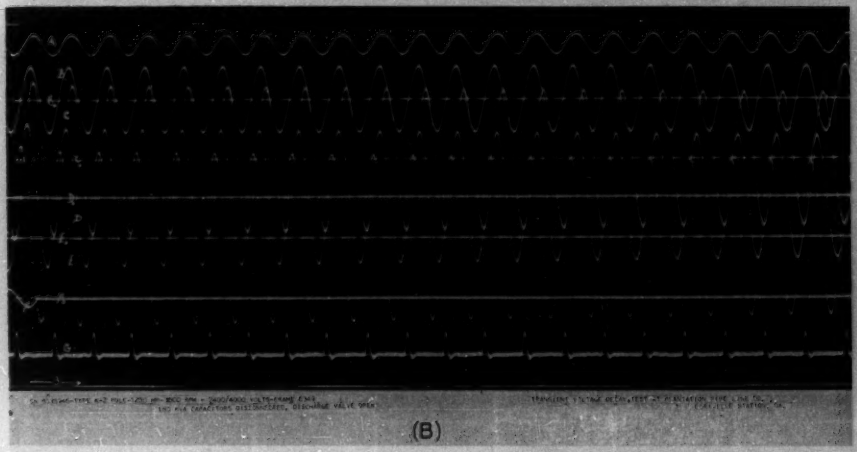


Fig. 8. Composite oscillogram of transient voltage decay test. Curve identification is as follows:

A—60-cycle timing wave
B—Bus voltage (L-L)
C—Motor voltage (L-L)
D—Motor voltage (L-L)
E—Motor voltage (L-L)
F—Motor current
G—Pip indicates one revolution



A—Power-factor corrective capacitors connected and discharge valve open



B—Power-factor corrective capacitors disconnected and discharge valve open

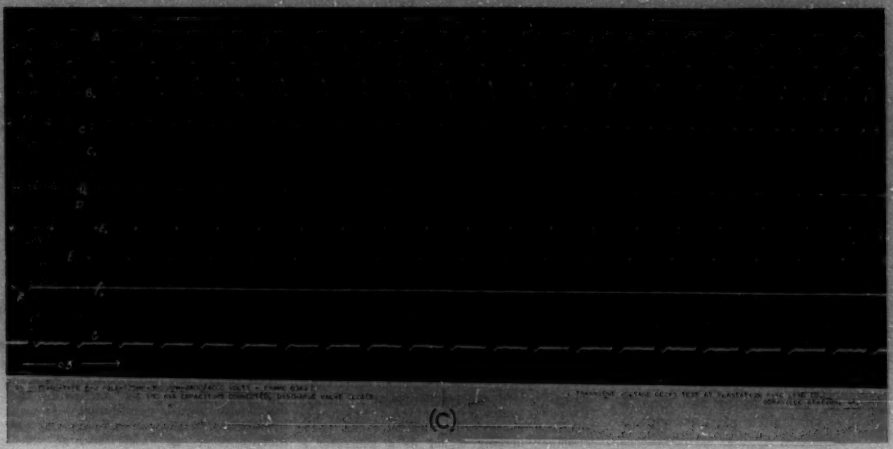
Fig. 10. Oscillograms for 20-cycle time interval from power interruption (see Fig. 8 subcaption for identification of curves). See facing page for Figs. 10(C) and (D)

may cause more rapid decay of residual voltage magnitude and frequency.[3] Under certain transmission-line and system-loading conditions, it is also possible to obtain self-excitation of the induction motors, producing rising voltage phenomena;[4] see Appendix III.

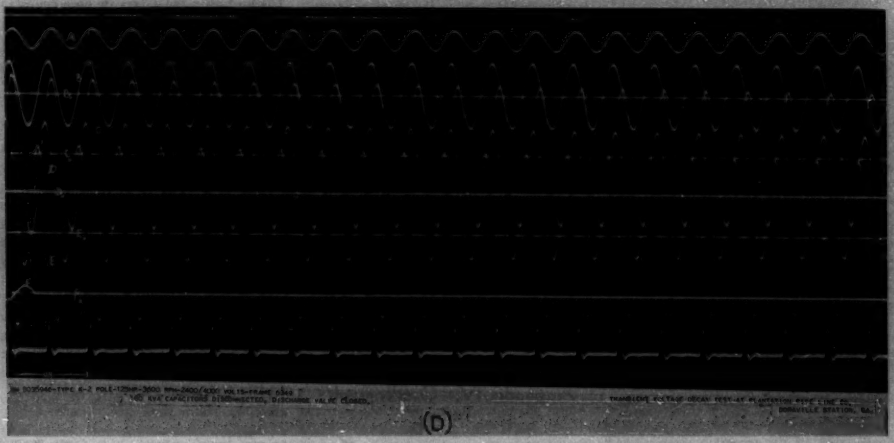
Field experience indicates that the practical considerations of power system loading, configuration, and machine design reduce the effects which could be encountered. However, the problem is present and can occur particularly where relatively long transmission lines feed isolated 1- or 2-unit pumping stations. Since each case requires individual evaluation, accurate generalizations are not practical.

A general solution to the problem could lie in the transmission-line relaying. Either of the following alternates appears desirable:

1. Dead-line check relay to ensure that no objectionable residual voltage remains



C—Power-factor corrective capacitors connected and discharge valve closed



D—Power-factor corrective capacitors disconnected and discharge valve closed

Fig. 10 (cont.).

on the load side of the line breaker.

2. Automatic reclosing on a fixed-time basis.

In the similar situation of transferring power station auxiliaries from one bus to another, the criterion is residual voltage 25% rated or less. Fig. 8 is a composite oscillograph indicating that the time delay between the opening of the power source until the residual decays to 25% of rated may be as long as 5.5 seconds.

Conclusions

Proper protection of pipe-line stations against transient disturbances consists of the following:

1. Station-type lightning arresters at transmission voltage level for the transformer supplying the station.

2. Surge protection on the station bus, consisting of a station-type lightning arrester and surge capacitor in parallel, located a minimum distance from the bus connection.

3. Power supply system having either:
   a. Dead-line check relaying.
   b. Time-delay automatic reclosure.

Appendix I. Effect of Multiple Outgoing Circuits

The effect of multiple outgoing circuits in reducing the magnitude of reflected surge voltages is illustrated by the following table:

Number of Outgoing Circuits	E_2, Voltage on Outgoing Circuits in Terms of Incoming Voltage E_1
0	$2E_1$
1	$1E_1$
2	$0.666E_1$
3	$0.5E_1$
4	$0.4E_1$
5	$0.333E_1$
6	$0.286E_1$

This is established by the relationship

$$E_2 = 2E_1 \frac{Z_2}{Z_1 + Z_2}$$

where

E_1 = incoming voltage
E_2 = outgoing (refracted) voltage
Z_1 = surge impedance of incoming line
Z_2 = surge impedance of outgoing line

Appendix II. Protection, Traveling Waves

For a traveling wave, localized stresses can be produced in the windings, the greatest stress occurring in the line terminal coils, the coils adjacent to the neutral end of Y-connected windings, and the mid-point between terminals on delta-connected windings. Due to the charging current of the capacitance between turns and from turns to slots, this charging current causes stress to exist on the line terminal coils. The distribution of this voltage is not uniform, which is the reason for the line coils having the highest induced voltage.

For the condition where a traveling wave enters the machine, on one of the three phases, the maximum neutral voltage will be approximately two thirds of the incoming line voltage due to the phenomenon of parallel surge impedances. If the incoming traveling wave enters on all three phases, the neutral voltage will be approximately double that of the incoming wave. The same condition exists for the delta-connected machine, except that the doubling of voltage occurs at the mid-point of the winding.

Due to the steepness of the wavefront, a voltage occurs between adjacent turns, the maximum value of which depends upon the steepness of this wavefront and the electrical length of the turns involved. To limit the turn-to-turn voltage stresses to an acceptable value, it is necessary to control the steepness of the wave entering the machine. This is done by placing a capacitance to ground at the terminals of the machine. In order to be most effective, the length of the connection from the line to the capacitor must always be shorter than the distance from the connecting point to the motor terminals. For maximum effect, the capacitor connection length should be kept to a minimum.

Appendix III. Motor Transients

Momentary power interruptions can produce undesirable transient conditions in operating motors. The motor which is operating has considerable stored energy, both mechanical energy stored in inertia of the motor and load, and electromagnetic energy stored in the magnetic field, the mechanical energy predominating. When the power source is interrupted, the electromagnetic energy continues to excite the machine, making it act as a generator. The effect is a residual voltage of a declining frequency related to the motor deceleration. The magnitude of voltage and its phase relation to the power system when power is reapplied determine the resulting transient conditions.

Fig. 4 shows a plot of the expected peak torque developed in the motor against the phase displacement of the motor residual voltage from the power system, in degrees. The torque which appears on the motor shaft can be calculated by multiplying the peak-torque value, read from Fig. 4, by the ratio of H_L (inertia constant of the load) to H_T (inertia constant of motor rotor and load). For example, if it is found that the load inertia constant is 25% of that of the motor rotor, then the torque appearing on the shaft will be 0.25/1.25 or 0.20 times that read from the curve.

The possibility of obtaining self-excitation and rising voltage phenomena is discussed in reference 4. This effect would be responsible for motor voltages in excess of rated being encountered, as in curves *1* and *2* of Fig. 4.

Appendix IV. Oscillograph Tests

Oscillograph tests were made (Fig. 9) for the following four conditions:

1. Suction and discharge valves open with power-factor corrective capacitors connected.
2. Suction and discharge valves open with capacitors disconnected.
3. Suction valve open and discharge valve closed with capacitors connected.
4. Suction valve open and discharge valve closed with capacitors disconnected.

These tests serve to evaluate, in addition to the basic power reapplication problem, the effect of the following:

1. Power-factor corrective capacitors.
2. Oil flow through the pump case.

Effect of Power-Factor Corrective Capacitors

The use of power-factor corrective capacitors, although they affect the magnitude of residual voltage, does not appreciable affect the phase shift, since this is related to the deceleration of the motor (inertia constant).

Effect of Motor Load

One of the peculiarities of this type of application is the "load" carried by the machine during the deceleration. As seen by referring to the curves taken with the suction valve open and discharge valve closed, the machine decelerates at a less rapid rate than with the valves open. It should be mentioned as a matter of interest that the increment is not appreciable. The load when the valves are open consists of the effect of reverse flow of fluid through the pump casing on the impeller. When power is removed from the motor, the difference in pressure between the pump suction pressure and the pump discharge pressure must equalize. Consequently, the flow reverses, and in flowing from discharge to suction exerts a decelerating torque on the motor. The effect noted is the difference in deceleration of the motor with the valves open and with the discharge valve closed. For example, at the end of 3 seconds, speed of the pump is 78% rated with the valves open and 84% rated with the valves closed. During the first 20 cycles this effect is negligible. Figs. 10(A) through (D) are oscillographic tests, under various conditions, for the 20-cycle time interval from power interruption.

References

1. Protection of Transmission Systems Against Lightning (book), W. W. Lewis. John Wiley & Sons, Inc., New York, N. Y., 1949, pp. 316–61.

2. Transfer of Steam-Electric Generating-Station Auxiliary Buses, D. G. Lewis, W. D. March. *AIEE Transactions*, vol. 74, pt. III, June 1955, pp. 322–34.

3. Analysis of Series Capacitor Applications Problems, J. W. Butler, C. Concordia. *Ibid.* (*Electrical Engineering*), vol. 56, Aug. 1937, pp. 975–88.

4. Transients in Power Systems (book), H. A. Peterson. John Wiley & Sons, Inc., New York, N. Y., 1951, pp. 126–83.

Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function

W. A. STEIN
MEMBER AIEE

G. J. THALER
MEMBER AIEE

Synopsis: This paper presents two general concepts which are very valuable in analyzing nonlinear servomechanisms. A graphical procedure is introduced by which the frequency response of a nonlinear closed-loop system can be determined from the describing function. The theory is valid for the general describing function which is both amplitude and frequency dependent. Second, an analytical method is developed for evaluating the describing function by Fourier analysis of that type of nonlinear component whose time constant and gain vary with the magnitude of the input signal. As an example of a nonlinear element which meets these specifications the 2-phase servomotor was selected. The closed-loop frequency response was then predicted by the graphical method and verified experimentally.

A METHOD is herein developed for obtaining the frequency response of a closed-loop nonlinear servo system from the describing function of the nonlinear element and the transfer function of the linear elements. In the literature[1] a procedure is presented for obtaining the frequency response of a closed-loop system from the describing function. However, this analysis is limited to special describing functions which are dependent on signal amplitude alone, such as saturation, and even then the method is quite involved.

The total transfer function G of a servomechanism can be resolved into the product of two components: $N(j\omega,V)$ which is the describing function of the nonlinear element, and $H(j\omega)$ which is the transfer function of the linear elements. A plot of $H(j\omega)$ and $-1/N(j\omega,V)$ is then made on a gain versus phase angle co-ordinate graph. The $-1/N(j\omega,V)$ curve is then the locus of the critical points.

These curves may then be used with a Nichols' chart to compute the frequency response of the system when the test signal has a fixed input amplitude R. The procedure involves some trial and error, and is as follows: Select any frequency, such as $\omega=\omega_3$. This frequency is to be used on both the $H(j\omega)$ and $-1/N(j\omega,V)$ curves. Since the $-1/N(j\omega,V)$ curves contain a number of points at frequency ω_3 (due to amplitude sensitivity), an intelligent estimate of V must be made and the corresponding ω_3 point located.

Assume $V=V'$. The point on the $-1/N$ plot corresponding to ω_3 and V' is pinpointed over the origin of a conventional Nichols' chart. A value of closed-loop gain M and a value of closed-loop phase shift ϕ are then read off the Nichols' chart contour lines at the point where $H(j\omega)$ is equal to ω_3.

It is then necessary to check these results against the assumed value $V=V'$. If a check is not obtained the procedure is repeated until it is, but convergence is usually very rapid. The method of checking is based on the following concepts:

M, the closed-loop gain, is defined as $|C/R|$ and ϕ is the angle between the output C and the input R in the closed-loop system. According to the foregoing definitions

$$C=MR\underline{/\phi} \tag{1}$$

The error E is then

$$E=R-C=R(1-M\underline{/\phi}) \tag{2}$$

The gain of the linear elements located between the error detector and the nonlinear element will be labeled $I(j\omega)$. Fig. 1 shows a servo system with a 2-phase motor representing the nonlinear element N.

In Fig. 1, $H=G_rG_cG_eG_g$ and I is $G_gG_eG_a$ or G_cG_a if the input signal is simulated by a dynamic analyzer.

If the estimate of V' is correct then

$$V'=I(j\omega)]_{\omega=\omega_3}E=I(j\omega)]_{\omega=\omega_3}R(1-M\underline{/\phi}) \tag{3}$$

otherwise a better value of V must be selected and the procedure repeated.

The entire closed-loop frequency response for a specific input amplitude R can be obtained by repeating these calculations at a number of frequencies.

In Fig. 2, the negative reciprocal of the describing function and the transfer function of the remaining linear elements are sketched on the gain versus phase angle chart. A Nichols' chart is superimposed at the critical point ω_3 and V'.

Analytical Method for Determining the Describing Function of Nonlinear Components in which the Nonlinearity Is Both Amplitude and Frequency Sensitive

By definition the describing function of any nonlinear device is a complex output/input ratio, in which the input is a pure sinusoid and the output is the fundamental component of the Fourier series representing the output wave produced by the sinusoidal input. The first step in the evaluation of a describing function is to write a general differential equation which describes the output quantity in terms of the input. Such an equation might be

$$\ddot{\theta}_o+f(\omega,V)\,\dot{\theta}_o+F(\omega,V)\,\theta_o+k_1\theta_o=k_2\theta_i \tag{4}$$

where

θ_i = the input to the nonlinear element
θ_o = the output of the nonlinear element

The next step is to evaluate the specific nonlinear coefficients, $f(\omega,V)$ and $F(\omega,V)$ which are, in general, functions of both frequency and input amplitude. This evaluation is often possible only in the form of experimental curves. The third step is to select an empirical equation for $f(\omega,V)$ and $F(\omega,V)$. In theory the choice of these equations is arbitrary, but in practice some consideration must be given to the requirements of ensuing manipulations.

A sinusoidal input is then specified

$$\theta_i=V\sin\omega t$$

and a steady-state solution assumed for the lowest derivative of the dependent variable in terms of a Fourier series with unknown coefficients, thus

$$\theta_o=A_1\sin\omega t+B_1\cos\omega t+A_2\sin 2\omega t+B_2\cos 2\omega t+A_3\sin 3\omega t+B_3\cos 3\omega t+\ldots \tag{5}$$

Paper 58-84, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N.Y., February 2-7, 1958. Manuscript submitted July 8, 1957; made available for printing November 21, 1957.

W. A. STEIN and G. J. THALER are with the U.S. Naval Postgraduate School, Monterey, Calif.

The authors acknowledge the assistance of the Office of Naval Research in obtaining special equipment for this project.

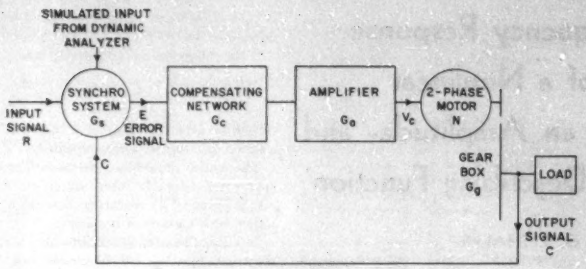


Fig. 1. Block diagram of servomechanism

Equations for higher derivatives are obtained by successive differentiations. All of these expressions for θ_i, θ_o, the derivatives of θ_o, $f(\omega,V)$, and $F(\omega,V)$ are substituted in the differential equation. A set of simultaneous equations is obtained from this equation by the process of equating corresponding trigonometric terms. These equations are then solved to determine A_1 and B_1, the coefficients of the fundamental terms in the output wave. The describing function is then:

$$N(j\omega,V)=\frac{\sqrt{A_1^2+B_1^2}}{V}\Big/\tan^{-1}\frac{B_1}{A_1} \qquad (6)$$

Application of Analytical Method in Determining Describing Function of a Specific Nonlinear Element, the 2-Phase Motor

The literature[2,3,4] discusses in detail the describing functions of simple nonlinearities which are only amplitude dependent such as backlash, coulomb friction, saturation, and hysteresis. Very little is done with nonlinearities which are both amplitude and frequency dependent.

As an example of a more complex nonlinear element, the 2-phase servomotor has been selected. Kochenburger[5] discusses velocity limiting in servomotors and acknowledges that such devices as the 2-phase motor and the hydraulic motor are highly nonlinear even apart from the velocity limiting effects that might be involved.

Both the gain and time constant are functions of the amplitude of the control winding voltage V_c as indicated in Fig. 3.

It is not a difficult undertaking to determine experimentally the gain in radians per second per volt and the time constant of a 2-phase servomotor. This procedure is adequately described in the literature.[6]

An examination of the speed versus torque characteristics of the 2-phase servomotor as obtained from dynamometer tests show that over the normal operating range the characteristics can be represented by straight lines. See Fig. 4.

The family of curves can be represented by the linear equation

$$T=AV_c-\left|\frac{\partial T}{\partial n}\right|\dot{\theta} \qquad (7)$$

where

$$A=\frac{\text{stalled torque at rated voltage}}{\text{rated voltage}}$$

V_c = rms value of voltage on the control field

$\left|\frac{\partial T}{\partial n}\right|$ = reciprocal of the slope of the speed-torque curves

θ = motor shaft position in radians

By assuming that all the developed torque goes into acceleration and overcoming viscous friction then

$$KV_c=\tau\ddot{\theta}+\dot{\theta} \qquad (8)$$

The gain constant is then

$$K=\frac{A}{F+\left|\frac{\partial T}{\partial n}\right|} \qquad (9)$$

and the time constant is

$$\tau=\frac{J}{F+\left|\frac{\partial T}{\partial n}\right|} \qquad (10)$$

It should be noted that τ is proportional to K, or more specifically

$$\tau=\frac{J}{A}K \qquad (11)$$

Both τ and K (which are functions of V_c) must be represented by empirical formulas in order to derive a describing function. After preliminary investigations it was decided to place the following restrictions on these empirical expressions:

1. The empirical expression which represents the K curve must (when multiplied by a constant term) also be a good approximation to the τ curve.

2. The expression must be an even function of V_c.

3. The denominator of these expressions must be capable of representation by a rapidly converging series.

Such a pair of expressions is

$$K=\frac{G}{\cosh FV_c} \qquad (12)$$

$$\tau=\frac{E}{\cosh FV_c} \qquad (13)$$

The curve of motor gain K versus V_c can usually be measured with sufficient accuracy in the laboratory; hence the G and F constants are evaluated by standard curve fitting methods to fit the motor gain curve best. The time constant τ versus V_c curve usually presents more difficulty in instrumentation, especially if the system is a fast one with small value of τ. Only the single constant E is evaluated in fitting the second expression to the τ versus V_c curve. The closeness of fit of the empirical expressions is shown in Fig. 3, except at very low values of V_c where stiction interferes with the measurements.

Theoretically the describing function of motor output position is obtained by exciting the motor with a sinusoidal signal in the control field of $v_c=V_c\sin\omega t$ and determining the fundamental component of amplitude of the motor shaft displacement in radians. The magnitude of the describing function is this fundamental component divided by V_c, and the angle of the describing function is the angular displacement between the input signal v_c and the fundamental of the motor shaft position. The 400-cycle-per-second sinusoidally modulated suppressed carrier wave which is the input to the control field is shown in Fig. 5.

V_c is the peak amplitude of the exciting signal measured in rms volts. In other words, if the rotating synchro which is producing the sinusoidally modulated suppressed carrier signal was stopped at a time when its output voltage was a maximum, the reading of an rms voltmeter would be V_c volts. The ω represents the angular speed of the rotating synchro in radians per second.

To continue the analytical derivation of the describing function of the 2-phase motor, substitute the empirical expressions of equations 12 and 13 into equa-

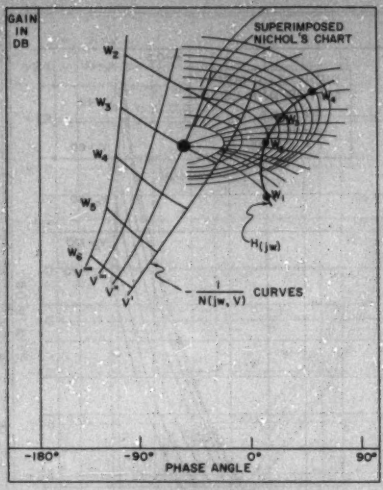


Fig. 2. Graphical construction for determining points on the frequency response characteristic

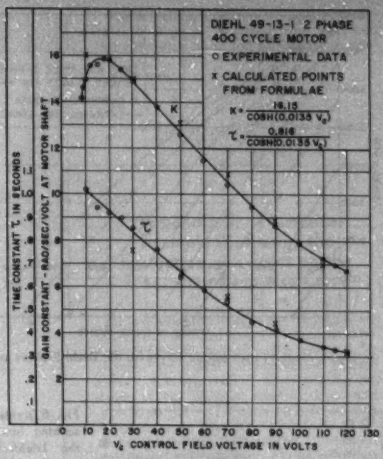

DIEHL 49-13-1 2 PHASE 400 CYCLE MOTOR
○ EXPERIMENTAL DATA
x CALCULATED POINTS FROM FORMULAE
K = 16.15 / COSH(0.0135 Vc)
τ = 0.816 / COSH(0.0135 Vc)
TIME CONSTANT τ IN SECONDS
GAIN CONSTANT - RAD/SEC/VOLT AT MOTOR SHAFT
K
τ
0 10 20 30 40 50 60 70 80 90 100 110 120 130
Vc CONTROL FIELD VOLTAGE IN VOLTS

Fig. 3. Two-phase motor characteristics: gain versus control field voltage and time constant versus control field voltage

tion 8. The input will be the sinusoidal signal, $V_e \sin \omega t$, instead of a constant rms voltage.

$$GV_e \sin \omega t = \theta E + \dot{\theta} \cosh (FV_e \sin \omega t) \quad (14)$$

The steady-state solution of equation 14 is assumed to be of the form:

$$\theta = A_1 \sin \omega t + B_1 \cos \omega t + A_3 \sin 3\omega t + B_3 \cos 3\omega t + \ldots \quad (15)$$

Note that in this case it is convenient to assume the Fourier series solution for the first derivative of the output.

Evaluation of the coefficients A_1 and B_1 define the fundamental component of the velocity response of the motor. Integration of this result then specifies the fundamental component of the output position, which is defined to be the describing function. Equations for evaluating the describing function $N(j\omega, V_e)$ which is a function of both angular frequency and amplitude of the motor input signal are derived in Appendix I.

Although the actual evaluation of a describing function plot might appear quite formidable, much of the data will be repetitious. The general plot of the describing function in the operating range of the motor can be accurately sketched by evaluating about ten well chosen points.

A plot of the describing function of shaft position of a Diehl motor with its associated load is shown in Fig. 6. It will be noted that the describing function is dependent on both the amplitude of the control field voltage and the modulating frequency ω.

The servomechanism of Fig. 1 without the compensating filter was assembled in the laboratory. The 2-phase servomotor was connected to its load through a high-precision 20-to-1 gear box having less than 1/4 of one degree of blacklash to insure that any nonlinearities present would be due to the motor alone. The amplifier was specially designed to have negligible saturation and phase shift. A tachometer was coupled to the motor as part of its normal load and the velocity output wave recorded for sinusoidal inputs. Fourier analyses were made of the output velocity wave for half a dozen combinations of input voltage and modulating frequencies ω. From these the describing function of motor velocity was evaluated and compared with corresponding points of the derived describing function. The discrepancies in amplitude averaged about 4% and in angles about 2 deg (degrees).

Obtaining the Frequency Response Characteristics From the Describing Function Plot

In order to verify the method of obtaining the frequency response characteristics outlined previously, the simple servomechanism without a compensating filter was tested in the laboratory. The transfer function of the linear components $H(j\omega)$ consisted of the product of the amplifier and synchro gain divided by the gear ratio and in this simple example was a single point for all frequencies. The negative reciprocal of the describing function $-1/N(j\omega, V_e)$ is plotted in Fig. 7. $H(j\omega)$ is a single point as shown.

The mathematical procedure for evaluating the frequency response at $\omega = 5$ radians per second is described in Appendix II.

Fig. 8 shows the frequency response obtained for this system with an input R of 0.69 rms volt. The crosses and circles represent points obtained by conventional testing on this same system.

The system in this example was deliberately loaded with a large inertia to produce low resonant frequency and a large motor time constant. This was done to facilitate measuring techniques. A Kearfott *111-2AB* motor with a 50-to-1 gear box was also tested. For this motor the parameters E, G, and F were evaluated to be 0.0492, 13.9, and 0.015 respectively. It was found that the value of M at resonance, the M_p, and the resonant frequency of these simple circuits varied with the magnitude of the input signal R.[7] The exact variation depended on the location of the H point. In general the

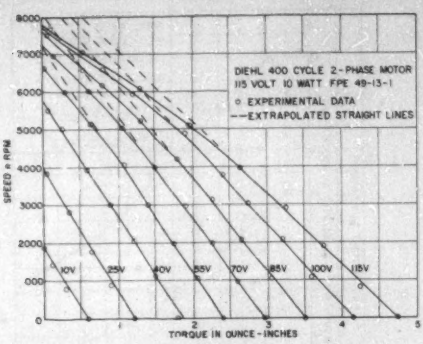


Fig. 4. Speed versus torque curves for the Diehl 2-phase motor

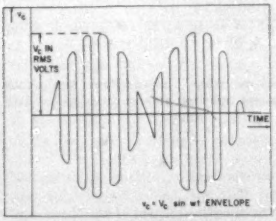


Fig. 5. Sinusoidally modulated suppressed carrier input wave

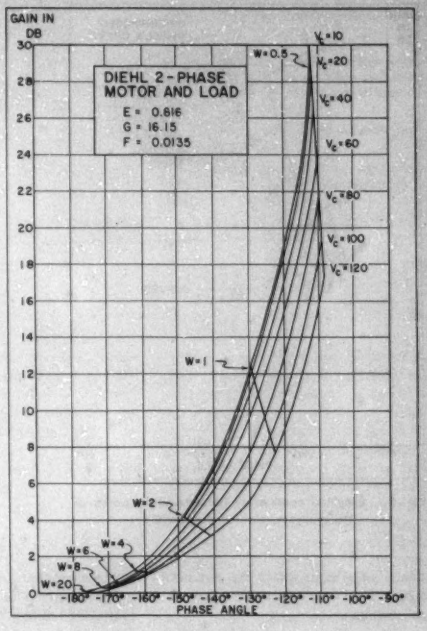


Fig. 6 (right). Describing function of the Diehl 2-phase motor with inertia load

resonant frequency increased as the amplitude of the input signal was decreased.

Additional Comments

The describing function of an instrument servomechanism obtained for a specific inertia load can be used even if the inertia load is altered. This is accomplished by changing the numerical values of ω plotted in Figs. 6 and 7 by the factor J_2/J_1 where J_1 is the original value of total inertia of the motor and load and J_2 is the new value.

By assuming the time constant τ as being invariant the method of obtaining the describing function of motor velocity will produce essentially the same describing function that Kochenburger gives for velocity limiting.

Conclusions

1. A method has been developed which permits computation of the frequency response of a nonlinear servomechanism when excited with a constant amplitude sinusoidal input of adjustable frequency.

2. The method is applicable not only when the nonlinearity is amplitude sensitive, but also when the nonlinearity is both amplitude and frequency sensitive.

3. A procedure for the derivation of describing functions has been developed. This procedure is specifically intended for cases where the nonlinearity is both amplitude and frequency sensitive.

4. Application of these techniques to a servo with nonlinear 2-phase motor resulted in an experimental confirmation of the theory.

5. These methods also provide useful information about signal levels at the input of the nonlinear component.

Appendix I. Derivation of the Describing Function of Motor Position Formulas

The differential equation describing the operation of the motor is

$$GV_c \sin \omega t = \ddot{\theta}E + \dot{\theta} \cosh (FV_c \sin \omega t) \quad (16)$$

The steady-state solution is then assumed to be

$$\dot{\theta} = A_1 \sin \omega t + B_1 \cos \omega t + A_3 \sin 3\omega t + B_3 \cos 3\omega t \quad (17)$$

Due to the symmetry of the motor there should be no even harmonics.

The time derivative of equation 17 is

$$\ddot{\theta} = \omega A_1 \cos \omega t - \omega B_1 \sin \omega t + 3\omega A_3 \cos 3\omega t - 3\omega B_3 \sin 3\omega t \quad (18)$$

Equations 17 and 18 are then substituted into the differential equation 16. The cosh $(FV_c \sin \omega t)$ is then expanded according to the trigonometric identity

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \ldots \quad (19)$$

By further trigonometric substitutions all the terms are reduced to expressions which are simply a constant multiplied by $\sin \omega t$, $\cos \omega t$, $\sin 3\omega t$, or $\cos 3\omega t$.

The coefficients of the $\sin \omega t$, $\cos \omega t$, $\sin 3\omega t$, and $\cos 3\omega t$ on both sides of the transformed equation are then equated to yield four simultaneous equations which are functions of A_1, B_1, A_3, B_3, E, F, G, V_c, and ω.

These four equations can then be solved for A_1, B_1, A_3, and B_3. Since the describing function method deals only with the fundamental of the wave, A_3 and B_3 are not required. Too large a relative magnitude for A_3 and B_3 indicate that the describing function method would not produce satisfactory results.

The solutions for A_1 and B_1 are

$$A_1 = \frac{\alpha}{\Delta} \quad (20)$$

$$B_1 = \frac{\beta}{\Delta} \quad (21)$$

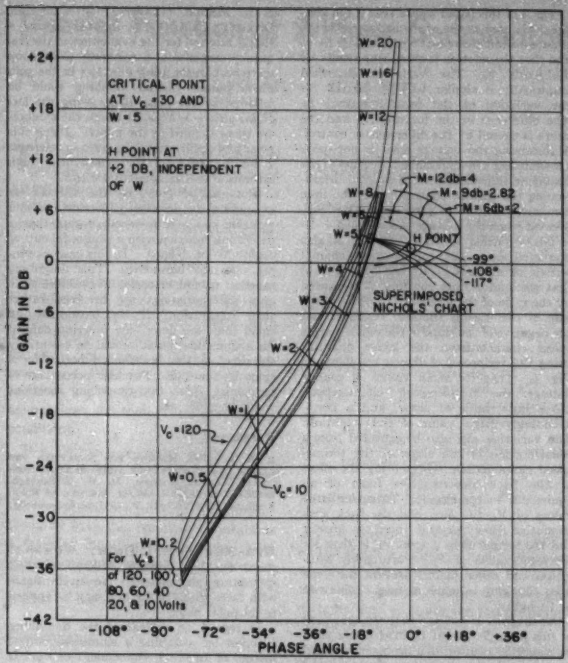


Fig. 7. Curves for calculating the frequency response of the 2-phase motor servo

where

$$\alpha = GV_c\left(1+\frac{5}{8}F^2V_c^2+\frac{7}{48}F^4V^4+9\ \omega^2E^2+\frac{9}{8}\omega^2E^2F^2V_c^2+\frac{9}{192}\omega^2E^2F^4V_c^4\right) \quad (22)$$

$$\beta = -GV_c\left(\omega E+\frac{1}{2}\omega E F^2V_c^2+\frac{27}{192}\omega EF^4V_c^4+9\ \omega^3E^3\right) \quad (23)$$

$$\Delta = 1+F^2V_c^2+\frac{75}{192}F^4V_c^4+10\omega^2E^2+9\ \omega^4E^4+5\ \omega^2E^2F^2V_c^2+\frac{171}{192}\omega^2E^2F^4V_c^4 \quad (24)$$

The describing function of motor velocity is

$$\frac{\sqrt{A_1^2+B_1^2}\ /-\tan^{-1}|B_1/A_1|}{V_c} \quad (25)$$

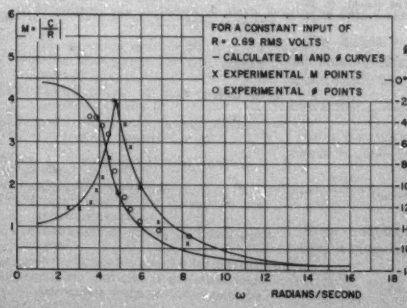


Fig. 8. Comparison of experimental and calculated frequency response curves

The describing function of motor position which is $N(j\omega, V_c)$ is found from equation 25 by dividing by ω and shifting the angle -90 deg.

Therefore:

$$N(j\omega, V_c) = \frac{\sqrt{A_1^2+B_1^2}\ /-90\text{ deg}-\tan^{-1}|B_1/A_1|}{\omega\ V_c} \quad (26)$$

Appendix II. Numerical Example Illustrating the Evaluation of a Point on the Frequency Response Curve

It is desired to evaluate the frequency response of the closed-loop system at an arbitrary value of ω, say $\omega=5$ radians per second. As a first approximation assume that $V_c=30$ rms volts. Upon placing the origin of the Nichols' chart at $\omega=5$ and $V_c=30$ on the $-1/N$ curve, it is found that $H(j\omega)$ falls on the $M=11$ decibels (3.54) and the $\phi=-108$-deg contours. The input amplitude R is 0.69 rms volt, so according to equation 2

$$E=0.69(1-3.54\ /-108\text{ deg})$$
$$=2.73\ /58.2\text{ deg}$$

Since the input inserted in the circuit is an electric signal simulated by a dynamic analyzer instead of a mechanical input, and this simple system has no series compensation, $I(j\omega)$ is merely the amplifier gain $G_a=9.7$.

According to equation 3

$$V_c=9.7\times2.73=26.5\text{ rms volts}$$

which is probably within reading accuracy of the initial assumption of $V_c=30$.

As a result, for an input of $R=0.69$ rms volt at $\omega=5$ radians per second, the point on the frequency response curve is $M=3.54$ and $\phi=-108$ deg.

References

1. Some Saturation Phenomena In Servomechanisms With Emphasis On The Tachometer Stabilized System, E. Levinson. *AIEE Transactions*, vol. 72, pt. II, Mar. 1953, pp. 1–9.

2. Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities, H. D. Grelf. *Ibid.*, Sept., pp. 243–48.

3. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, R. J. Kochenburger. *Ibid.*, vol. 69, pt. II, 1950, pp. 270–84.

4. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements, E. C. Johnson. *Ibid.*, vol. 71, pt. II, July 1952, pp. 169–81.

5. Limiting in Feedback Control Systems, R. J. Kochenburger. *Ibid.*, vol. 72, pt. II, July 1953, pp. 180–94.

6. Evaluating the Effect of Nonlinearity in a 2-Phase Servomotor, W. A. Stein, G. J. Thaler. *Ibid.*, vol. 73, 1954 (Jan. 1955 section), pp. 518–21.

7. A Generalized Method for Determining the Closed-Loop Frequency Response of Nonlinear Systems, L. T. Prince. *Ibid.*, Sept. 1954, pp. 217–24.

(Discussion follows)

Discussion

G. H. Fett (University of Illinois, Urbana, Ill.): This paper presents an interesting method for solving the problem of obtaining the gain-frequency response of a nonlinear servomechanism for a constant magnitude of input signal. The fine check between experimental and calculated values shown in Fig. 8 attests to the care and accuracy of the work. However, it should be pointed out that the example given is one for which the peak value of M is about 4, and it is just for such lively systems that the method may be applied.

The paper also provides an important confirmation of the theoretical relationship describing the variation of the corner frequencies of the 2-phase induction motor under load given in one paper at the same session.[1]

In equation 25 of the paper referred to the corner frequencies are given by

$$s_{1,2}=(T_e+T_m)/2T_eT_m\pm \{[(T_e+T_m)/2T_eT_m]^2-K(I^2+I_c^2)/T_eT_m\}^{1/2}$$

where T_e and T_m are the electrical time constant of the motor winding and the mechanical time constant of the motor and load respectively, K is a motor constant, I is the constant current in one phase, and I_c is the controlled current in the other phase of the 2-phase induction motor. Now it is known that $T_e \ll T_m$, it is assumed that $I^2>I_c^2$, and since the corner frequencies found in practice are real and widely different,[2]

$$(1/T_e+1/T_m)^2-4KI^2/T_eT_m>4KI_c^2/T_eT_m$$

Then write

$$x^2=K(I_c/I)^2/T_eT_m[(1/2T_eI)^2-K/T_eT_m]$$

and so

$$s_{1,2}=1/2T_e\pm[1/2T_e-I(K/T_eT_m)^{1/2}]\times(1-x^2)^{1/2}$$

From this relation for the corner frequencies it can be seen that for the corner frequency of the smallest magnitude the value will increase as the value of x increases. Further, if $I(K/T_eT_m)^{1/2}$ is neglected compared with $1/2T_e$, then

$$s\approx(1/2T_e)[1-(1-x^2)^{1/2}]$$

The time constant is the reciprocal of the corner frequency and so the principal time constant of the motor and load is

$$T=2T_e/[1-(1-x^2)^{1/2}]$$

Fig. 3 of this paper shows that if the voltage V_c is used as the controlled variable the time constant decreases with increase in V_c and that the relation is given empirically by $T\alpha 1/\cosh V_c$. The variation described empirically is similar to that derived for the variation of the control current I_c. The difference in the formulation and the curve is caused by the difference in control. Maintaining the voltage fixed is not quite the same as a constant current because of the change in impedance of the circuit. Maintaining the current control makes the time constant sensitive to current which is suppressed when the voltage is controlled.

The variation in time constant is also predicted from the torque-speed characteristic of the motor. It has been shown[3] that the motor time constant is measured by the ratio of the moment of inertia to the sum of the viscous friction coefficient and the negative of the slope of the motor torque speed characteristic, the latter quantity being the reciprocal of the curves shown in Fig. 4. Thus for small values of control voltage the torque-speed characteristic has a small value of slope, and a correspondingly large value of time constant. The variation can also be expected from a consideration of the slope of the torque-speed characteristic representing the effect of the back electromotive force of an equivalent motor circuit. Thus, for large values of V_c, the flux, and the back electromotive force per-unit speed, is greater and the torque drop is more rapid than for the small values of V_c. Fortunately saturation and other factors prevent the slope from changing as much as might otherwise be thought.

Although it is realized that the emphasis in this paper is on the method rather than on a specific solution to a problem, it would be interesting to see the results of the application of a number of signals of different amplitude, similar to the single one of Fig. 8. Of especial interest would be the observation of the shift in the resonant frequency of the system with the change of amplitude, such as would be expected on the basis of the shift in the time constant with V_c. Have the authors such curves?

References

1. Transfer Function of 2-Phase Servomotors, S. L. Mikhail, G. H. Fett. *AIEE Transactions*, vol. 77, pt. II, May 1958, pp. 97–99.
2. Transfer Function for a 2-Phase Induction Servo Motor, Lloyd O. Brown, Jr. *Ibid.*, vol. 70, pt. II, 1951, pp. 1890–93.
3. Feedback Control Systems (book), G. H. Fett. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1954, pp. 36, 268.

S. L. Mikhail (University of California, Berkeley, Calif.): The authors present a simple method for the evaluation of the frequency response of a nonlinear closed-loop servomechanism using dual loci in the gain phase plane. Similar thinking could be followed in the polar plane using the loci $H(j\omega)$ and $-1/N(j\omega, V)$ where the symbols are those defined in the paper. Using this polar plot applications are given in reference 1 for a positional servomechanism with saturation and relay nonlinearities.

The method given by the authors for evaluating the describing function is also a valuable one. It is worth noting that a describing function with a sinusoidal output could also be defined. In this case, the input contains harmonics. This describing function should be easier to evaluate since only differentiations are involved rather than integration. If the harmonics are small then the describing function defined for a sinusoidal input should be essentially the same as the describing function for a sinusoidal output. For high percentage of harmonics these two describing functions would be different.

Reference

1. A General Method for Analyzing the Closed Loop Response of a Linear and a Nonlinear Servomechanism, H. H. El-Sabbagh. *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., 1957, pt. 4, pp. 58–77.

W. A. Stein and G. J. Thaler: We wish to thank Dr. Fett and Dr. Mikhail for their interesting discussions. We quite agree with them that the method may be applied to the polar plane.

In regard to evaluating the describing function by assuming a sinusoidal output instead of an input, the amount of work is the same in the numerical type of analysis as made in the paper. Actually there is no integration involved either way.

In regard to the example which has a peak M of 4, a poorly damped system was deliberately selected to emphasize the accuracy of the method. The authors do not know of any limitations on the application of the method.

In answer to the question concerning the shift in resonant frequency with the amplitude of input signal, it was found that both the amplitude of M_p and the resonant frequency shifted. In general the resonant frequency increased as the amplitude of the input signal was decreased. The manner in which M_p varied for changing amplitude of input depended on the location of the $H(j\omega)$ point.

Transfer Function of 2-Phase Servomotors

S. L. MIKHAIL
MEMBER AIEE

G. H. FETT
MEMBER AIEE

THE extensive use of 2-phase induction motors in servo systems has justified giving some attention to their analysis under transient conditions. Stanley[1] gives stator and rotor differential equations referred to axes fixed in the stator. Neal,[2] Hraba,[3] and Brown[4] give an experimental investigation of the problem in addition to approximating the nonlinear equations to linear ones with approximate transfer functions in the last two references. Hopkin,[5] Koopman,[6] and Gilfillan and Kaplan[7] give solutions to the problem under certain assumptions and under certain operating conditions.

It is the purpose of this paper to give the general form of the torque-speed equation for the motor if the currents in the two phases are given as functions of time and from this equation to establish a transfer function relating the speed or angular displacement of the motor and the input signal for a modulated wave applied to one of the two phases.

Assumptions

The following assumptions are made:

1. The input stator currents are known as functions of time as in the case of a current source for both phases or the stator currents are assumed to vary linearly with the stator terminal voltages.

2. The effects of saturation, hysteresis, and eddy currents are neglected.

3. Sinusoidal distribution of turns is assumed.

Nomenclature

A = gain given by equation 24

$B(s) = \dfrac{G(s)}{\Theta(s)}$ = displacement transfer function given by equation 23

F = the frictional torque per unit angular velocity

$f_1(t)$, $f_2(t)$ = functions of time defined by equations 14 and 15

$g(t)$ = modulating function in control phase

$G(s)$ = the Laplace transform of $g(t)$

$H(s) = \dfrac{G(s)}{\Theta^*(s)}$ = velocity transfer function given by equation 22

I = the rms value of the current in the reference phase

i_1, i_2 = instantaneous values of stator currents

i_a, i_b = instantaneous values of rotor currents

i_{aa}, i_{ab} = rotor currents giving same magnetomotive force as stator currents

J = the moment of inertia of the rotor

K = torque constant introduced in equation 3

L = inductance of a rotor winding

M = maximum value of mutual inductance between a stator and a rotor winding

$p = d/dt$

r = the resistance of a rotor winding

T_e = electric torque

T_m = mechanical torque

θ = instantaneous value of angle between rotor and stator winding

$\theta^* = p\theta$

$\theta^{**} = p^2\theta$

$\Theta(s)$ = the Laplace transform of θ

$\Theta^*(s)$ = the Laplace transform of θ^*

$\tau_e = L/r$ is the rotor electrical time constant

$\tau_m = J/F$ is the rotor mechanical time constant

$\omega = 2\pi f$, where f is the frequency of the reference current

Fundamental Equations

With reference to Fig. 1, let *1* and *2* be the stator windings, *a* and *b* the rotor windings, respectively; thus, i_a, i_b are the instantaneous values of the rotor currents, and i_1, i_2 are the instantaneous values of the stator currents.

If i_{aa} and i_{ab} are defined as components of rotor currents which give the same magnetomotive force as the stator currents, then,

$$i_{aa} = i_1 \cos\theta + i_2 \sin\theta \tag{1}$$

and

$$i_{ab} = i_2 \cos\theta - i_1 \sin\theta \tag{2}$$

Since the rotor windings, are closed, the differential equations for the rotor windings are given by

$$0 = (r+Lp)i_a + Mpi_{aa} \tag{3}$$

and

$$0 = (r+Lp)i_b + Mpi_{ab} \tag{4}$$

The electric torque is proportional to the vector product of the currents in the *a* and *b* axes and is thus given by

$$T_e = KM[i_a i_{ab} - i_b i_{aa}] \tag{5}$$

From equations 3, 4, and 5, the electric torque is given by

$$T_e = \frac{KM^2}{(r+Lp)}\,[i_{aa}pi_{ab} - i_{ab}pi_{aa}] \tag{6}$$

From equations 1, 2, and 6 the electric torque is given by

$$T_e = \frac{KM^2}{(r+Lp)}\,[(i_1pi_2 - i_2pi_1) - (i_1^2+i_2^2)\theta^*] \tag{7}$$

If $\tau_e = L/r$ is the rotor electrical time constant, equation 7 becomes

$$T_e = \frac{KM^2}{r(1+\tau_e p)}\,[(i_1pi_2 - i_2pi_1) - (i_1^2+i_2^2)\theta^*] \tag{8}$$

A special case would occur when i_1 and i_2 are sine waves of the same frequency and amplitude and displaced 90 degrees apart in time or if

$$i_1 = I_m \cos\omega t$$

and

$$i_2 = I_m \sin\omega t$$

The expression for the torque becomes

$$T_e = \frac{KM^2}{r(1+\tau_e p)}\,I_m^2(\omega - \theta^*) \tag{9}$$

Equation 9 shows that in the case of balanced stator currents, the electrical steady-state torque is proportional to the square of the current, to the slip, and to the cube of the mutual inductance, and is inversely proportional to the rotor resistance.

For any general variation of i_1 and i_2, however, equation 8 gives the desired electric torque. To get the relation between θ^*, i_1, and i_2, the electric torque T_e is equated to the mechanical torque T_m which is given by

$$T_m = F\theta^* + J\theta^{**} \tag{10}$$

Now if $\tau_m = J/F$ is the mechanical rotor time constant, the mechanical torque is given by

$$T_m = F(1+\tau_m p)\theta^* \tag{11}$$

Equating equations 8 and 11 gives

$$\left[\frac{Fr}{KM^2}(1+\tau_m p)(1+\tau_e p) + (i_1^2+i_2^2)\right]\theta^* = i_1pi_2 - i_2pi_1 \tag{12}$$

Equation 12 represents the general linear differential equation with variable coefficients relating θ^*, with i_1 and i_2 which are known functions of time. The solution of equation 12 by approximate methods or by computer techniques would be the solution to the problem.

If τ_e is neglected compared to τ_m, equation 12 gives

$$p\theta^* + \left[\frac{1}{\tau_m} + \frac{KM^2}{Fr\tau_m}(i_1^2+i_2^2)\right]\theta^* = \frac{KM^2}{Fr\tau_m}(i_1pi_2 - i_2pi_1) \tag{13}$$

Paper 58-12, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted August 8, 1957; made available for printing October 23, 1957.

S. L. MIKHAIL is with the University of California, Berkeley, Calif., and G. H. FETT is with the University of Illinois, Urbana, Ill.

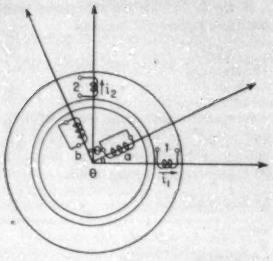


Fig. 1. Stator and rotor windings

If

$$f_1(t) = \frac{1}{\tau_m} + \frac{KM^2}{Fr\tau_m}(i_1^2 + i_2^2) \tag{14}$$

and

$$f_2(t) = \frac{KM^2}{Fr\tau_m}(i_1 pi_2 - i_2 pi_1) \tag{15}$$

equation 13 gives

$$p\theta^* + f_1(t)\theta^* = f_2(t) \tag{16}$$

The solution to this equation is of the form[6]

$$\theta^* = e^{-\int f_1(t)dt} \int f_2(t)\, e^{\int f_1(t)dt} dt \tag{17}$$

Transfer Function for a Modulated Wave

Let i_1 be the reference current, thus

$$i_1 = \sqrt{2}\, I \cos \omega t \tag{18}$$

and

$$i_2 = \sqrt{2} g(t) \sin \omega t \tag{19}$$

Equation 12 thus gives equation 20:

$$\left[\frac{Fr}{KM^2}(1+\tau_m p)(1+\tau_e p) + 2I^2 \cos^2 \omega t + 2g^2(t) \sin^2 \omega t\right]\theta^* = 2\omega I g(t) + Ig'(t) \sin 2\omega t \tag{20}$$

Assuming the variation of $g(t)$ with time to be small relative to the reference, $g'(t)$ is small and the term on the right-hand side representing twice the reference frequency can be neglected. Also, neglecting terms on the left-hand side producing harmonic generation between θ^* and $g(t)$, there is

$$\left[\frac{Fr}{KM^2}(1+\tau_m p)(1+\tau_e p) + (I^2+g^2)\right]\theta^* = 2\omega I g(t) \tag{21}$$

where g is the rms value of $g(t)$. The transfer function relating the Laplace transform of g and θ^* is thus given by

$$H(s) = \frac{G(s)}{\Theta^*(s)} = \frac{2\omega I}{\frac{Fr}{KM^2}[(1+\tau_m s)(1+\tau_e s)] + (I^2+g^2)} = \frac{A}{(s+s_1)(s+s_2)} \tag{22}$$

and the transfer function relating the Laplace transforms of g and the displacement θ is given by

$$B(s) = \frac{G(s)}{\Theta(s)} = \frac{A}{s(s+s_1)(s+s_2)} \tag{23}$$

where

$$A = \frac{2\omega KM^2 I}{Fr\tau_m\tau_e} \tag{24}$$

$$s_{1,2} = -\frac{1}{2}\left(\frac{1}{\tau_e} + \frac{1}{\tau_m}\right) \pm \sqrt{\left[\frac{1}{2}\left(\frac{1}{\tau_e} + \frac{1}{\tau_m}\right)\right]^2 - \frac{KM^2(I^2+g^2)}{Fr\tau_m\tau_e}} \tag{25}$$

Equation 25 shows that the transfer function for the 2-phase induction motor has two poles at s_1 and s_2 which always lie in the left-hand plane.

s_1, s_2 are real when

$$\left[\frac{1}{2}\left(\frac{1}{\tau_e} + \frac{1}{\tau_m}\right)\right]^2 \geqq \frac{KM^2(I^2+g^2)}{Fr\tau_m\tau_e} \tag{26}$$

Realizing that τ_e is small relative to τ_m, then s_1, s_2 are real when

$$\frac{\tau_m}{\tau_e} \geqq 2 + \frac{4KM^2}{Fr}(I^2+g^2) \tag{27}$$

When this condition is fulfilled, s_1 and s_2 are the two corner frequencies[4] in the plot of transfer function amplitude against frequency using logarithmic scales. If, however, the condition given by equation 27 is not fulfilled, s_1 and s_2 are complex and the transfer function plot is determined by the damping ratio of these roots.

References

1. An Analysis of the Induction Motor, H. C. Stanley. *AIEE Transactions*, vol. 57, 1938, pp. 751–55.

2. Unbalanced Voltage Operation of 2-Phase Induction Motor, J. F. Neal. *Thesis*, University of Illinois, Urbana, Ill., 1950.

3. Transient Speed Response of 2-Phase Servomotors in Suppressed Carrier Systems, John Hraba. *Ibid.*, 1955.

4. Transfer Function for a 2-Phase Induction Servo Motor, Lloyd O. Brown, Jr. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1890–93.

5. Transient Response of Small-Phase Induction Motors, A. M. Hopkin. *Ibid.*, pt. I, pp. 881–86.

6. Operating Characteristics of 2-Phase Servomotors, R. J. W. Koopman. *Ibid.*, vol. 68, 1949, pp. 319–27.

7. Transient Torques in Squirrel-Cage Induction Motors, With Special Reference to Plugging, E. S. Gilfillan, E. L. Kaplan. *Ibid.*, vol. 60, 1941, pp. 1200–09.

8. Differential Equations (book), H. B. Phillips. John Wiley & Sons, Inc., New York, N. Y., 1922.

Discussion

Andris Padegs (Carnegie Institute of Technology, Pittsburgh, Pa.): In some of the previous analytical approaches to the response of 2-phase servomotors[1,2] solutions for low frequencies of modulation have been obtained by neglecting the effects of rotor currents on stator currents (or by assuming constant-current power supplies) and by neglecting the effects of rotor self-inductance. These approximations yield results according to which there is no phase shift between the generated torque and the envelope of the signal applied to the control winding in the case of 90-degree phase displacement between the carriers of the applied signals and according to which the torque-speed characteristics are linear.

The foregoing results do not agree with experimental observations, and there is evidence of additional phase shift due to the electrical parameters of the motor. By employing the Heaviside operational calculus for the solution of the differential equations, the authors of the paper have included the effect of rotor self-inductance, and, as expected, their solution contains also an "electrical" time constant. The mathematical treatment of the problem, however, is not rigorous, and this raises questions as to the validity of the conclusions.

The operational calculus as developed by Heaviside, permitting the use of the operator p as an algebraic quantity, is applicable to the solution of certain simple integrodifferential equations, but the original method did not have a rigorous mathematical justification. The validity of the Heaviside operational calculus has been subsequently justified by integration in the complex plane, and fragmentary conditions of validity are obtained by analogy with the Laplace transformation method. The authors of the paper, in fact, use the Heaviside operational method synonymously with the Laplace transform, as indicated by the replacement of the operator p by the Laplace variable s when going from equation 21 to equation 22. It follows then that an identical result should be obtained by using the Laplace transform throughout the analysis.

The Laplace transform, however, has precise conditions of validity, and, according to the real multiplication theorem, the transform of a product of two functions is not equal to the product of the Laplace transforms of the two functions, but has to be evaluated by complex convolution.[8] As given by equation 5, the torque is proportional to the difference between the products of two time functions (currents). In order

to obtain equation 6, the authors express the currents i_a and i_b in terms of the operator p and of the currents i_{oa} and i_{ob}, respectively, and substitute the resultant expression into equation 5. When using the Laplace transform, the analogous operation would not be valid, and hence there is no justification for it in the method used in the paper. The Heaviside operational method does not have an interpretation for the multiplication of two functions of time.

The paper also does not clearly emphasize the existence of a carrier frequency and the fact that the input intelligence appears in terms of the side bands of the control winding signal. It is reasonable to expect that the motor torque as given by equation 8 depends upon the relative magnitudes of data and carrier frequencies, and that the value of the carrier frequency might present some limitations on the bandwidth of the motor. In deriving equations 8 and 12 the authors have not stated over what frequency range the results are valid. Since no assumptions about the relative magnitudes of data and carrier frequencies have been made, this implies that the value of the carrier frequency is not important and that it does not set any limitations on the magnitude of the data frequency.

When sinusoidal modulated control-winding signals are considered, even under the assumption of constant-current power supplies, exact expressions for the steady-state torque as a function of the frequency of modulation are rather complicated since in general the torque depends upon motor speed, upon the phase and magnitude of the control winding current, and upon the relative magnitudes of the modulating and carrier frequencies. In the case of 90-degree phase displacement between the carriers of the input currents, which is the case considered in the paper and which is of most practical importance, the phase shift at stall between the generated torque and the envelope of the modulated control winding current can be approximated by a motor electrical time constant τ_e given by

$$\tau_e \approx \frac{LR}{R^2 + \omega_c^2 L^2}$$

where L is the rotor self-inductance, R is the rotor resistance, and ω_c is the carrier angular frequency. Thus the rotor electrical time constant has to be modified by a factor $R^2/(R^2 + \omega_c^2 L^2)$. This approximation is fairly close up to frequencies equal to one half of the carrier frequency. At higher modulating frequencies and under different operating conditions the phase lag may be appreciably different and may even exceed 90 degrees. However, when considering the application of the motor at frequencies of the order of carrier frequency, other factors affecting the bandwidth of the motor have to be included, particularly the effect of side-band torques. The factors affecting the bandwidth of the 2-phase servomotor have been investigated by this discusser, and it is planned to publish the results in the near future.

References

1. Carrier Frequency Servomechanisms. A. Sobczyk. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 246, July 1948, pp. 21–43.

2. 2-Phase Servo Motor Operation for Varying Phase Angle of the Control Winding Applied Voltage, Mark A. Steinbacker, Wilbur Meserve. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1987–93.

3. Transients in Linear Systems (book), M. F. Gardner, J. L. Barnes. John Wiley & Sons, Inc., New York, N. Y., 1942.

An Electro-Hydraulic Speed and Load Division Control for Constant-Speed Air-Turbine Drives

P. DANTOWITZ
NONMEMBER AIEE

L. G. NORRIS
NONMEMBER AIEE

Synopsis: This paper presents a description and analysis of an electro-hydraulic speed and load division control for an aircraft constant-speed air-turbine drive for a 400-cps (cycles per second) 60-kva alternator. The development program was undertaken as a product improvement effort over an existing mechanical speed control system. At this writing the development program has been completed and the system is undergoing flight and qualification testing. Representative performance data, in the form of test results, are presented.

THE DEMAND for high-quality constant-frequency electric power for aircraft electric systems has been increasing rapidly over the past decade. This demand has, in turn, placed increasingly stringent requirements on the electric system and drive controls. In order to satisfy these requirements, an electro-hydraulic speed control development program was initiated. An electric system approach was chosen for the following most significant reasons:

1. The greater immediate potential for higher accuracy of control as demonstrated by an earlier breadboard-type mechanization.

2. The significant improvements which had taken place in the reliability of electric components and in methods of manufacture and packaging to achieve physically rugged assemblies.

3. The relative ease of handling intelligence, that is, of adding and subtracting signals and of shaping frequency response characteristics.

4. Ease of adjustment and modification during the development program.

5. Ease of instrumentation and measurement. The ability to measure the behavior of the control system at a number of points in the control loop quickly and accurately is a tremendous advantage in the development testing of the prototype system.

Paper 58-138, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted March 1, 1957; made available for printing December 16, 1957.

P. Dantowitz and L. G. Norris are with the General Electric Company, West Lynn, Mass.

Statement of the Problem

The air-turbine drive which was to form the application for the new control system was in existence at the time the program was initiated. The drive control system at that time comprised a mechanical governor and a hydraulic actuation system for controlling the position of a throttle valve which, in turn, controlled engine compressor bleed air flow to the turbine.

The open-loop block diagram of the drive itself, without controls, is essentially as shown in Fig. 1. The hydraulic actuator and control valve for controlling air flow to the turbine is represented as an integration having a gain K_1, the Laplace transform notation being used. The input to this block is control oil flow and the output is in the form of an aerodynamic torque on the turbine. The gain K_1 is subject to relatively wide variations in magnitude depending upon the engine compressor bleed energy condition available at the turbine.

An unusual feature of the speed-control system of an air-turbine drive using jet-engine compressor bleed air is that it is really many systems. With reference to Fig. 1, the gain K_1 of the integration between control oil flow and turbine torque is a variable, depending primarily upon compressor discharge pressure, turbine exhaust pressure (airplane altitude), and

throttle valve position. The synthesis of such a control system consists of specifying one set of control system parameters which will result in satisfactory operation at the extremes of operating conditions. This system was designed to meet three such extreme operating conditions. These are set forth in detail along with the design approach philosophy in the section on "System Synthesis" which describes the system synthesis technique.

The drive and alternator forming the application for the subject control system development is rated at 60 kva, 0.75 power factor, with a fault load requirement of approximately 200% rated load. The airplane application calls for a system of four drives and alternators operating in parallel with an automatic real load division control forming a part of the drive control system. The design objectives for the program were as follows:

Dynamic speed accuracy about average speed	±0.1%
Accuracy of average speed (steady state)	±0.5%
Dynamic load division accuracy about average load	±10% of rated load
Accuracy of average load division (steady state)	±10% of rated load

The dynamic load division accuracy which measures the stability with which load is shared is perhaps the most difficult requirement to meet. It places a demand for excellent low amplitude performance on the control system in two ways.

1. Any self-induced low-amplitude disturbances in the speed-control system are accentuated by the very lightly damped resonance which characterizes parallel operating synchronous generators in aircraft. Such self-induced disturbances must therefore be held to a minimum.
2. The speed-control system contributes a necessary amount of damping to the forementioned resonance. This damping is not present at amplitudes lower than the threshhold level of the control and the load division system will in general oscillate at this amplitude. To limit the load oscillations to an acceptable value requires a very low control system threshhold.

Description of System

After the initial analytical work was completed, the investigation of several possible systems was initiated utilizing an analog computer to control the actual drive over the range of operating conditions. This approach was found to be advantageous since it permitted the operation of several systems with the drive to be controlled, along with establishment of system and component tolerances, based on actual operating conditions, considerably in advance of the first mechanization of the system.

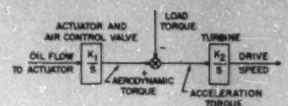


Fig. 1. Drive block diagram

The mechanization of the system which was finally chosen as a result of this approach is shown in Fig. 2(A). Drive speed is sensed by means of a permanent magnet generator driven through the turbine reduction gearing. The output of the permanent magnet generator is introduced into an *R-C* (resistance-capacitance) frequency discriminator, which provides the speed reference and from which a direct voltage proportional to the magnitude and direction of speed error is obtained. This error signal is then amplified by means of a magnetic amplifier, the output of which is a differential current. The magnetic amplifier load is the torque motor of a 2-stage electro-hydraulic servo valve. The servo valve responds to the electrical error signal to port control oil flow to one side or the other of the hydraulic actuator which controls the position of the air control valve.

Stabilization for the speed-control loop is provided by means of an internal feedback loop around the magnetic amplifier, the net effect of which is to generate a lead-lag in the forward speed-control loop. The magnetic amplifier feedback arrangement is also utilized to generate a very high static gain to overcome drifts and loading effects downstream of the amplifier with a relatively small steady-state frequency error. The stabilization circuitry is shown in Fig. 7(A), and its characteristics are treated in greater detail in the section on "Stabilization and High-Static Gain Techniques" of this paper.

The unit is started by opening the start-stop valve, the air-control valve being spring loaded to the open position. It will be appreciated that the primary speed control produces a valve opening signal during start until the reference speed is exceeded. Under most operating conditions, the turbine acceleration with only the primary speed control active would be very high and excessive speed overshoot would therefore result. A start circuit which is sensitive to turbine acceleration is therefore provided to control acceleration during start. This circuit is shown in Fig. 2(B), and it makes use of the rising voltage characteristic of the permanent magnet generator during start.

The output of the permanent magnet generator is rectified through a bridge rectifier and fed into a series *R-C* combination comprising R_1 and C_2. The output of the bridge is filtered by a shunt capacitor C_1 and the resistor R_3 provides a discharge path for the capacitor C_2.

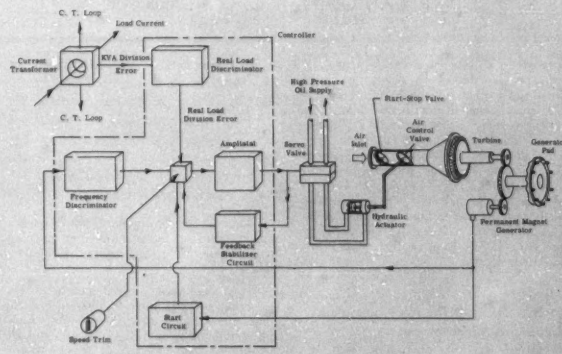


Fig. 2(A). Control system schematic

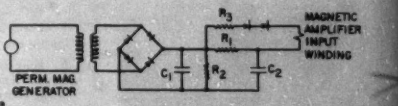


Fig. 2(B). Start circuit

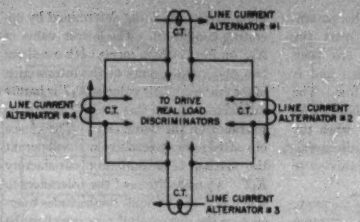


Fig. 3 (left). Current transformer loop

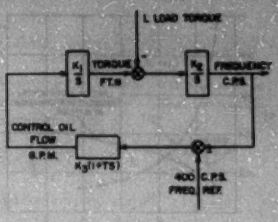


Fig. 4 (right). Simplified system block diagram

The input signal to the start winding of the magnetic amplifier is taken across R_1 and a resistor R_2 is utilized to provide the proper gain. The forward, low signal level characteristics of a pair of series diodes are utilized to provide a high impedance to relatively small signals so that the circuit is made almost completely in sensitive to steady-state leakage through the capacitor C_2 as well as to signal levels produced by normal system transients. The valve closing current i_c for a step input of drive acceleration is then approximately:

$$i_c = K_3 a \left(1 - e^{-\frac{t}{T}}\right)$$

where T is the R-C time constant, a is drive acceleration, and K_3 is the net gain. Then for any valve opening current i_0 produced by the primary control, the equilibrium acceleration a_e is approximately

$$a_e = \frac{i_0}{K_3\left(1 - e^{-\frac{t}{T}}\right)}$$

the final value being

$$a_e = \frac{i_0}{K_3}$$

It will be seen from the preceding that the equilibrium acceleration is a function of time and that the R-C time constant must therefore be considered. In other words, at $t=0$, the acceleration is essentially limited only by the drive and alternator inertias, and the time at which the proper closing current is obtained and the equilibrium acceleration is attained is a function not only of these parameters, but also of the R-C time constant.

Speed adjustment is obtained by means of a remotely positioned rheostat, which controls the d-c level of an electric signal to a separate winding of the magnetic amplifier. The speed trim is provided to permit exact speed synchronization for parallel operation.

Load division between paralleled drives is accomplished by means of a current transformer loop, shown in Fig. 3, in combination with a real load discriminator, which is a part of each drive control system. The output of the current transformer loop to each drive is a signal proportional to the difference between the average of all load currents and the load current from that drive's alternator. The real load discriminator is sensitive only to the component of the load error current in phase with the line voltage. The output of the real load discriminator is therefore an electric signal proportional to the real load division error on that drive, as compared to the average of all drive loads. The output of the real load discriminator is fed into another input winding of the magnetic amplifier and it acts on the speed-control loop in the same manner as speed error, controlling the position of the air control valve to maintain a balanced load.

It will be appreciated that the stability of the load division loop is sensitive to a number of parameters of the electrical system. In the particular application for which this system is designed, it was found necessary to attenuate the load error signal at higher frequencies because of a resonance. This is accomplished by means of a series lag incorporated in the real load dsicriminator output. The magnetic amplifier, the real load discriminator, the frequency discriminator, and the starting and stabilization circuits are all packaged in a single assemby identified as the controller.

System Synthesis

The system synthesis consists essentially of specifying the static and dynamic performance of the controller and servo valve with tolerances thereon so that the control will function satisfactorily at all flight conditions. Three conditions of operation were chosen such that the entire map of operating conditions would be covered if satisfactory performance could be obtained at these three.

The first condition is that resulting in maximum loop gain and occurs at maximum compressor discharge pressure, sea level exhaust, and a control valve position required for maximum electric load. The air pressure at the drive inlet is about 200 psia (pounds per square inch absolute) at this condition. The requirement here is to insure an adequate margin of stability. Phase margin at the crossover frequency of the open loop was taken as the stability criterion, a minimum of 30 degrees being the limit. The system parameter which identifies an operating condition is the gain of the throttle valve and it is shown as K_1 on Fig. 1. K_1 at this condition is 1,300 ft-lb/sec/gpm (foot-pounds/second/gallon per minute) and is the maximum encountered at any operating condition. The open-loop crossover frequency at this condition will therefore be the highest and limits on allowable high-frequency lags can be set to meet the criterion of 30-degree phase margin. In order to ease the component design problems, the crossover frequency at this condition should be kept as low as possible commensurate with the requirements at the other two extreme operating conditions.

The second design operating condition was chosen as the sudden application and removal of a fault load. The system requirement here is that the maximum deviation in frequency be held within satisfactory limits. Actually the sudden removal of the fault load is more important since the turbine has an overspeed tripout mechanism which shuts the unit down whenever the frequency exceeds 430 cps. The maximum frequency error was therefore set at 30 cps, steady state frequency being 400 cps.

A study of the operating map showed that at 55,000 feet altitude and engines idle, the travel of the throttle valve from no load to the fault load was at a maximum. If the maximum frequency upon removal of the fault can be met here, it will be met at any other operating condition.

Although the gain K_1 varies somewhat from the no load condition to the fault condition, an average gain of 190 ft-lb/-sec/gpm was chosen for synthesis purposes. The air pressure at the drive inlet

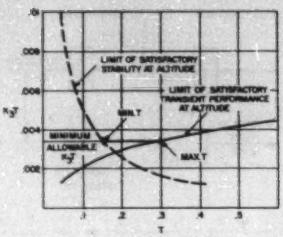


Fig. 5. Performance limits at extreme operating conditions

is 11.5 psia at this condition. The next step was to establish the form of the elements required to complete the loop. Because of the existence of the two integrators, some lead must be generated in closing the loop.

It should be noted here that the presence of appreciable speed-dependent machine torque would provide natural damping and alleviate the stabilization problem. In this application the speed-dependent torque varies with operating condition and actually reaches zero at one point. At its maximum value the speed-dependent machine torque results in a machine time constant of about 4 seconds, very close to an integration.

Fig. 4 shows a block diagram which completes the loop. Note that the gain of the servo valve has been lumped with the gain of the controller for convenience. A lead has been introduced to supply phase margin and it is intended that this be generated in the controller. The first step in the synthesis process is to fix K_3 which is the product of the servo valve and controller gain and T which is the controller lead break time constant.

The relationship between the parameters K_3 and T required to meet the transient requirement of 30 cps peak frequency error was determined and is shown on Fig. 5 as a solid line. The reason for the choice of the product TK_3 for this plot will become clear when the stability at high loop gain is discussed. Only values of TK_3 lying above this curve are satisfactory.

Having established the limits for meeting the second condition, let us examine the third operating condition. This condition is that resulting in minimum loop gain for a steady-state load (excludes fault loads) and occurs at 55,000 feet altitude, idle engines, and high pressure drop in the ducting leading from the engines to the drive. The air pressure at the drive inlet at this condition is 9 psia. The value of K_1 is 75 ft-lb/sec/gpm. As in the first condition, the requirement here is to maintain an adequate margin of stability and again 30-degree phase margin at open loop crossover frequency was chosen as the criterion. The relationship between K_3 and T required to meet 30 degrees phase margin was determined and plotted in Fig. 5 as a dotted line. Again only values of TK_3 lying above this curve are satisfactory.

We now return to the first condition where the loop gain is a maximum (K_1= 1,300 ft-lb/sec/gpm). The loop gain here is 18 times the gain at the third condition. Clearly at this gain condition, the crossover frequency will be higher than $1/T$. The crossover frequency of the straight line approximation at high gain is therefore

$$w_c = K_1 K_3 K_2 T$$

Since the design goal at this condition is to keep the crossover frequency as low as possible, we must set K_3T at a minimum, K_1 and K_2 being determined by operating conditions. The lowest value of K_3T on Fig. 5 which meets both condition two and three occurs at the intersection of the curves. This value of K_3T is hardly satisfactory since this choice allows no tolerance on the value of T. The choice of a satisfactory tolerance on T will permit the setting of a minimum satisfactory K_3T. Consideration of the tolerances to which K_3T can be held (this includes tolerance on the servo valve gain as well as the controller gain) determines the maximum value of K_3T. The maximum crossover frequency at high gain can now be determined and allowable phase lag or lead at this frequency assigned to each component to meet the 30-degree phase margin criterion.

In this case, the tolerance on T was chosen as ±33%. The value of K_3T at which this spread occurs is 0.0034, this being K_3T_{min}. Corresponding time constant T is 0.2 ±33% seconds. A tolerance of ±25% on K_3T was selected, requiring a maximum value of K_3T=1.65 times the minimum. The result is K_3T_{max}=0.0055. The maximum crossover frequency at high gain (K_1=1300 ft-lb/sec/gpm) is therefore 64 radians per second. The servo valve was allowed to contribute 10 degrees phase shift at this frequency. The lag due to the two integrators is 180 degrees. The minimum allowable phase lead of the controller at crossover is therefore 40 degrees, leaving a phase margin of 30 degrees. The tolerance on K_3T was split at ±10% on the gain of the servo valve and ±15% on the high frequency gain of the controller.

The next step in the synthesis process was to determine the nominal gain split between controller and servo valve. Because of the conditional stability of the system low amplitude sustained oscillations can be expected in the presence of certain low amplitude insensitivities. To minimize turbine speed oscillations resulting from the low amplitude insensitivity of the servo valve, its gain was as low as

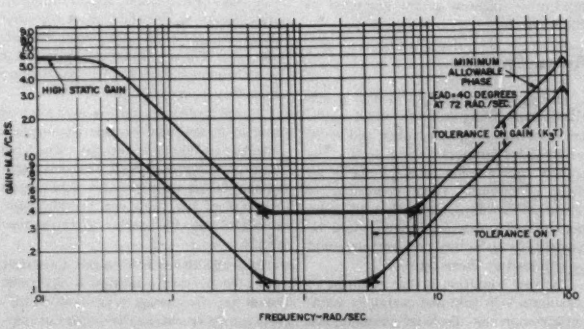


Fig. 6(A). Magnetic amplifier frequency response

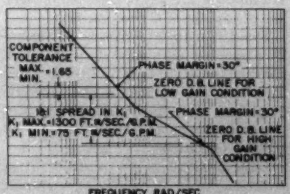

COMPONENT TOLERANCE MAX. +1.65 MIN.
PHASE MARGIN=30°
ZERO D.B. LINE FOR LOW GAIN CONDITION
±1 SPREAD IN K_1
K_1 MAX.=1300 FT. #/SEC./G.P.M.
K_1 MIN.=75 FT. #/SEC./G.P.M.
PHASE MARGIN=30°
ZERO D.B. LINE FOR HIGH GAIN CONDITION
FREQUENCY RAD./SEC.

Fig. 6(B). System bode plot

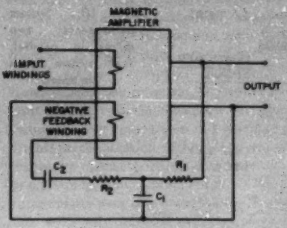


Fig. 7(A). Stabilization circuit

possible and the controller gain was thus established. Finally, to minimize steady-state frequency error resulting from servo valve null shift, the static gain of the controller was raised to approximately 15 times the gain determined by the foregoing synthesis. The gain is dropped down to substantially the latter value at 1/2 radian per second so that the effect of this low-frequency shaping is small at the crossover frequency of the system; see Fig. 6(A).

The synthesis technique used here is not exact; however, it served as an excellent starting point for the detail design of the controller and servo valve. As the design of these components progressed, more was learned about their characteristics. At the same time, more was being learned about the drive itself from actual test. The system analysis was continuously refined and required adjustments were made on the component specifications, until the final form was developed. Fig. 6(A) shows the final form of the controller. Bode plots of the system at maximum and minimum gain with phase margin at crossover as noted are shown on Fig. 6(B).

As a final check on the transient performance of the system with fault load removal at 55,000 feet idle engines, an analog computer study was conducted. This study showed that the maximum frequency error was held within the specified limit of 30 cps.

Stabilization and High-Static Gain Techniques

Stabilization Method

As was pointed out earlier, the presence of the two series integrations in the speed control loop requires that some form of stabilization be introduced. This is accomplished by negative feedback around the magnetic amplifier in the manner shown in Fig. 7(A). The block diagram and transfer functions for this arrangement are shown in Fig. 7(B), where T_1 is the

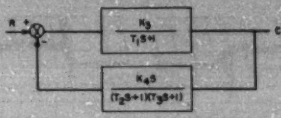


Fig. 7(B). Magnetic amplifier block diagram

time constant of the magnetic amplifier, T_2 is approximately $(R_1+R_2)C_2$, T_3 is approximately $R_1R_2C_1/R_1+R_2$, and K_3 and K_4 are the forward and feedback gains respectively.

In considering the analysis of the stabilizing circuitry, the very low frequency characteristics of the magnetic amplifier and feedback circuit may be neglected. The feedback transfer function may then be simplified to

$$\frac{K_3}{(T_3S+1)}$$

where

$$K_4=\frac{K_4}{T_2}$$

Then the closed-loop transfer function becomes

$$\frac{C}{R}(S)=\frac{K_3(T_3S+1)}{T_1T_3S^2+(T_1+T_3)S+(K_3K_4+1)}$$

Analyzing the quadratic term, it can be shown, where $K_3K_4>>1$, that the natural frequency W_n is

$$w_n=\sqrt{\frac{K_3K_4}{T_1T_3}}$$

and the damping ratio φ is

$$\varphi=\frac{T_1+T_3}{2\sqrt{K_3K_4T_1T_3}}$$

For the purpose of rapid synthesis using bode techniques, it was found convenient to consider an effective time constant T_4 defined by the relationship

$$T_4=\frac{T_3}{K_3K_4}$$

It will be observed that T_4 defines the frequency of the apparent intercept between the lead assymptote generated by T_3 and the static gain of the magnetic amplifier. Using this relationship, the characteristics of the quadratic become

$$w_n=\frac{1}{\sqrt{T_1T_4}}$$

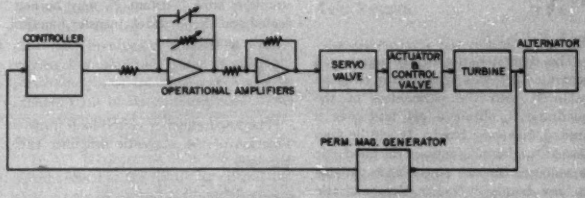


Fig. 8. Test setup for low amplitude performance

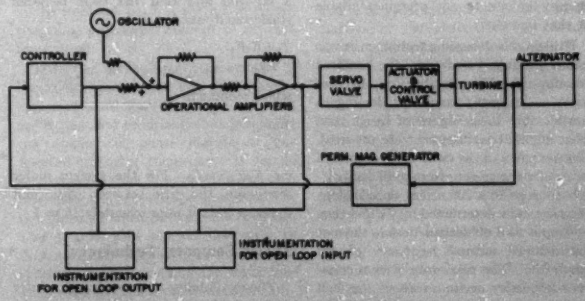


Fig. 9. Test setup for frequency response

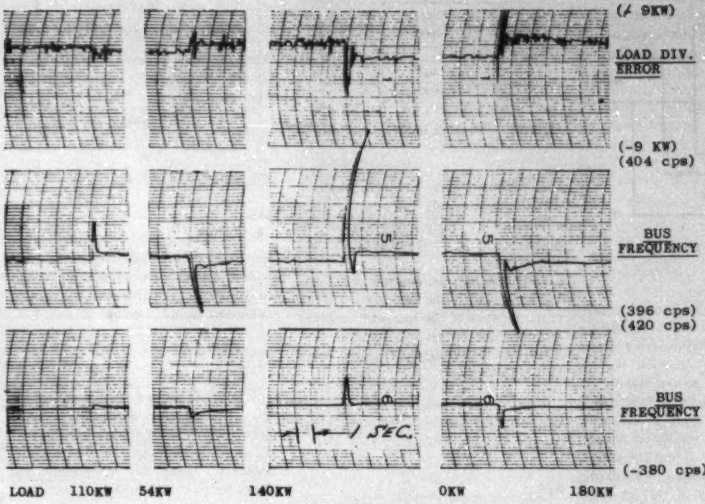


Fig. 10. System performance, two drives in parallel. Total system loads are noted

and

$$\varphi=\frac{1}{2}\sqrt{\frac{T_4}{T_1}}$$

for $K_3K_3T_4>>T_1$.

The characteristics of the series lead generated by this circuit can then be balanced with the parameters of the quadratic to obtain a net lead over a limited frequency band. It will be seen that as the damping ratio of the quadratic is reduced, the net phase lead available at any frequency below the quadratic natural frequency is increased. At the same time, however, the peak amplitude ratio at the quadratic natural frequency is increased an this effect must be balanced against the system gain margin available at that frequency.

Utilizing the foregoing technique, it was found possible to provide the required amount of lead for stabilization purposes with a magnetic amplifier having a considerably lower figure of merit than what might at first appear to be required. The net effect of the circuit just described is to raise the system from a 40 decibel/-decade slope to a 20 decibel/decade slope at a frequency determined by T_3, and then to drop it to a 60 decibel/decade slope at the damped natural frequency of the quadratic. The maximum system cross-over frequency occurs at about one half the damped natural frequency of the quadratic.

Provision of High Static Gain

At very low frequencies, the magnetic amplifier time constant T_1 may be neglected and the feedback transfer function may be simplified to

$$\frac{K_4S}{T_3s+1}$$

The low-frequency closed-loop transfer function of the magnetic amplifier then becomes

$$\frac{C}{R}(S)=\frac{K_4(T_3S+1)}{(T_3+K_3K_4)s+1}$$

It will be seen that the low-frequency closed-loop response is thus provided with a lag-lead and that the ratio between static and dynamic gain is

$$\frac{T_3+K_3K_4}{T_3}$$

which reduces to

$$\frac{K_3K_4}{T_3}$$

for $K_3K_4>>T_3$. For the system under discussion, the ratio between static and dynamic gain is approximately 15 to 1.

Analog Computer Techniques

The availability of an analog computer during both the initial design and the development phases of the program resulted in the generation of a number of development and evaluation techniques utilizing the computer. Several of these are presented here as being of general interest since they serve to demonstrate the tremendous versatility and potential of analog computers in programs of this nature.

As was pointed out earlier, the existence of drive hardware at the beginning of the control system development program permitted the operation of the drive itself, both in single unit and in parallel operation, with the control system analogued on a computer. In this manner, the suitability of the system design was checked out and the characteristics and tolerances on the various components verified.

As the program progressed into the mechanization stage, many variations of this initial approach were used and these permitted a rapid evolution of the system from the breadboard to the final configuration. Many combinations of physical components and analog-generated characteristics were employed, not only to permit gradual introduction of new elements into the system but also to allow immediate checks to be made on any given component or element suspected of causing faulty operation of the system.

For instance, the system was operated with only the magnetic amplifier and its feedback network simulated on the

computer in order to compare operation of the system with particular magnetic amplifier units in the loop and with their required characteristics accurately simulated on the computer. At various times during the program, the controller start circuit was simulated on the computer to allow comparison of a particular circuit mechanization with a set of known computer characteristics when operating in a given system configuration. Various components were also operated in the system with only selected portions of their total circuitry analogued. For instance, an early breadboard controller was operated in the drive control system with only the feedback circuit of its magnetic amplifier analogued on the computer. Utilizing these techniques, it was found that any component capable of being analogued could be checked out very rapidly by inserting its theoretically perfect characteristics in the system and comparing operating results.

It was also found possible to investigate such factors as loading effects and reflected impedances of one component on another by isolating the components through an analog computer circuit. Such a technique was utilized in investigating the effects of the servo valve static and dynamic loading on the magnetic amplifier. In that case, the loop was opened at the magnetic amplifier output and a pure, variable resistance load placed on the magnetic amplifier. The signal was then picked up from the load resistors and introduced into computer operational amplifiers, the output of which were connected to the servo valve. Using this arrangement, the load on the amplifier could be adjusted, and the gains and frequency response characteristics of the amplifiers adjusted and modified to permit comparison of operation over a wide range of conditions. A schematic of such an arrangement is shown in Fig. 8.

The analog computer was also found to be useful in generating special instrumentation circuitry. For instance, various types of filter circuits, or circuits having particular frequency response characteristics could be put together on the computer in a very short time. The characteristics of these circuits could be modified at any time with relative ease to suit particular requirements.

The analog computer was also utilized obtain the open-loop frequency response characteristics of the system with the system operating under closed-loop control. It will be appreciated that such a technique avoids the necessity for controlling the steady state level of the open loop while obtaining its characteristics.

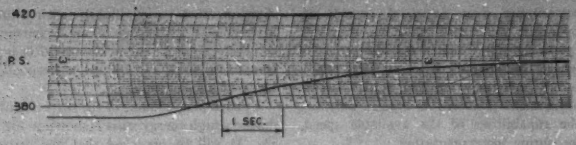


Fig. 11. Typical performance of the start circuit during a high-energy start

The manner in which this was accomplished is shown in Fig. 9.

In the particular case shown, the loop was opened at the controller output and a pair of computer operational amplifiers was inserted in series with the loop as shown. The first of these is a summing type amplifier, which permits the addition of a signal from an oscillator. The second amplifier is included in this case only to avoid a sign reversal.

With this arrangement, the closed loop may be oscillated at any desired frequency and amplitude. The net input to the open loop was measured at the servo valve input and the output was measured at the controller output. In addition to the instrumentation shown, the input and output of each component was instrumented so that the frequency response characteristics of the components while operating in the loop were also checked at the same time.

Another testing technique that was employed involves a method of determining low-amplitude gain margin and phase margin of the mechanized system including all of its nonlinearities. This was accomplished utilizing the arrangement shown in Fig. 8.

It will be observed that the loop was again opened at the controller output and that a pair of operational amplifiers was inserted in series with the loop. One of these amplifiers is provided with a variable feedback resistance to permit a gain adjustment as well as a variable capacitor in the feedback path to permit introduction of an adjustable amount of lag into the system.

With this arrangement, the amount of lag at any preselected loop gain can be increased until low amplitude sustained oscillations are produced, thus indicating the amount of phase margin at any selected crossover frequency. It should be recognized that, in setting the crossover frequency the attenuation introduced by the added lag must be taken into account. The system gain margin may also be determined by increasing the gain of the amplifier without adding lag until low-amplitude sustained oscillations are produced. The susceptibility of the system to various types of oscillations, such as those caused by the presence of significant nonlinearities, may be readily investigated with this technique.

Utilizing this testing technique, it was found that sustained oscillations were produced by the addition of approximately 30-degree lag at the maximum and minimum crossover frequencies of the system, indicating that the system, as finally mechanized, met the design phase margin requirements. It was also found that amplitude limited oscillations were encountered with only very slightly less lag added to the system than that which produced oscillations which rapidly increased in amplitude, indicating that the system did not contain any nonlinearities capable of producing sustained oscillations within the design range. The gain margin checks produced similar results.

Test Results

Representative test results, taken during qualification testing of the subject drive control system, are shown in Figs. 10 and 11.

It will be observed that in Fig. 10 bus frequency has been recorded simultaneously on two channels, the upper of the two having a range of 396–404 cps and being used primarily to observed the stability of the system about average speed, and the lower channel having a range of 380–420 cps to permit observation of transient performance. The system load at each point is set forth at the bottom of the chart and the changes in the system load are step changes.

The results shown in Fig. 10 are typical of the performance obtained with two drives operating in parallel. Ground testing of the drives in the airplane showed the performance to be substantially the same to slightly better for four drives operating in parallel. Typical speed accuracy in parallel was found to be $\pm 0.025\%$ (± 0.1 cps) about average speed, while dynamic load division about average load per machine was generally within $1 \pm$ kw over the rated load range.

Drift in the absolute value of average speed from the initial setting without

retrimming over sustained periods of operation was found to be less than 1.0 cps in those portions of the test program where this characteristic could be monitored. Accuracy of the initial speed adjustment primarily is limited only by the capabilities of the available speed measuring instrumentation.

Average load division drift from initial setting was found to be less than 2 kw although practical limitations in the test program did not permit accumulation of a great deal of data on this characteristic. The average load per machine in parallel may be adjusted, of course, by varying the setting of the trim rheostat. It will be observed that in the tests represented by the data of Fig. 10, the load division error varies slightly with total system load, increasing by an increment of about 2 kw as the system load is increased from 0 kw to 180 kw, corresponding to 90 kw per drive.

Typical performance of the start circuit during a high energy start is shown in Fig. 11. It will be observed that the drive acceleration is held sufficiently low by the start circuit during starting to avoid any overshoot of the steady-state speed level.

Conclusions

1. The results of the program indicated that a much higher level of performance could be obtained for the same or less development effort with an electrohydraulic control than could be realized with a mechanical-hydraulic system.

2. The approach of using an analog computer as an operating control to investigate various approaches and system tolerance limits early in the program was found to be of considerable advantage. When coupled with the relative speed with which the electric components may then be mechanized to definitely known requirements, it would appear that such an approach would offer similar advantages under other circumstances.

3. It was found that the electrical approach permitted greater ease and speed of evaluating the prototype system once it had been mechanized.

Classical Heat Flow Problems Applied to Induction Billet Heating

R. M. BAKER
MEMBER AIEE

THE AUTHOR has often referred to the solutions of certain classical heat-flow problems as a method of analyzing the more practical heat-flow problems of induction heating. Certainly others have done the same but surprisingly little has been published on the subject. Several difficulties confront the average engineer considering such an approach. In the first place, the mathematics used in the solution of classical heat flow problems is rather formidable and strange to many. Like anything else, the mathematics becomes understandable and appears simple enough when one masters it, but this requires time and effort and is a deterrent to the practicing engineer. An even greater deterrent is the fact that the numerical evaluation of the solution is often tedious or may even involve an arbitrary function which must be integrated for the special case of the problem at hand before numerical evaluation is possible. Two classical heat-flow problems are selected in this paper for evaluation and use in the analysis of induction billet heating. To overcome the difficulties cited, the solutions are reduced to simple formulas and dimensionless curves. The general plan of presentation is 1. to evaluate and extend the classical solutions to obtain working formulas and curves, and 2. to show how these may be used to analyze the relationships involved in induction billet heating. As usual, it will be necessary occasionally to introduce logical approximations in applying the classical solutions.

It was originally planned to base the analysis of billet heating on the classical solution to the problem of a cylinder heated with constant power input at the surface. In the process of re-evaluating this solution, it occurred to the author that an accurate solution to the problem of an induction-heated cylinder or billet (heat generated below the surface) could be obtained if one restriction was admitted. The solution indicates accurately the temperature distribution in the cylinder only after power has been applied for a certain specified time depending on the physical size and properties of the billet. Fortunately this restriction is no handicap in the present analysis and the use of this more accurate solution adds a certain elegance to the treatment which would have been lacking without it.

Paper 58-202, recommended by the AIEE Electric Heating Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted March 13, 1957; made available for printing December 2, 1957.

R. M. BAKER is with the Westinghouse Electric Corporation, Baltimore, Md.

In the classical solutions, as in their applications, it is convenient to have temperature curves start from zero. Thus any temperature curve starting from zero should be considered a curve of temperature rise above ambient.

Units

The equations of this paper may be used with any consistent system of units. However since electrical considerations are involved, it was found less confusing to adhere to centimeter-gram-second units throughout with temperature in degrees centigrade. Since electric power is usually expressed in watts and since power in this paper will be in calories per second (cal/sec) it is fitting to recall that 1 cal/sec = 4.19 watts. Symbols of quantities used and their units are listed in the following.

Nomenclature

v = temperature or temperature difference, degree centigrade (C)
t = time, seconds
a = radius of cylinder or billet, centimeters (cm)
r = variable radius, cm
P_0 = surface power density, $\frac{\text{cal}}{\text{sec}}/\text{cm}^2$
k_c = thermal conductivity, $\frac{\text{cal}}{\text{sec}}\,\text{cm}^{-1}\,\text{C}^{-1}$
c = specific heat $\frac{\text{cal}}{\text{gms (grams)} \times \text{C}}$
γ = density, gm/cm^3
$k = (k_c/\gamma c)$ = diffusivity, cm^2/sec
$T = kt/a^2$ = dimensionless time variable
V = a specific temperature or temperature difference, C
$\left.\begin{matrix}\text{ber}\ (x)\\ \text{bei}\ (x)\\ \text{ber}'\ (x)\\ \text{bei}'\ (x)\end{matrix}\right\}$ = Modified Bessel's functions and their derivatives

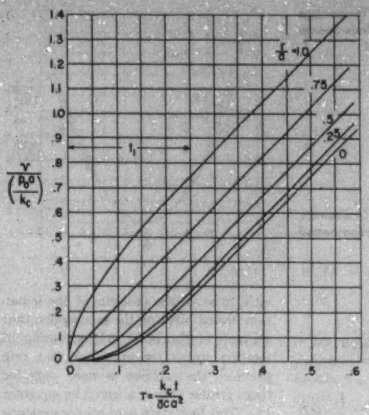
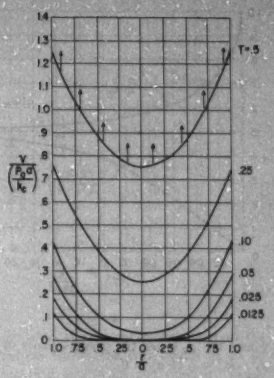


Fig. 1 (left). Temperature rise curves: constant power input at surface

Fig. 2 (right). Temperature distribution curves: constant power input at surface



f = frequency, cycles/sec
ρ = electrical resistivity, ohm-cm
μ = effective magnetic permeability, relative
k_1 = see equation 14
k_2 = see equation 14
δ = depth of current penetration, cm
P_a = average volume power density, $\frac{\text{cal}}{\text{sec}}/\text{cm}^3$
$X(x)$ = a function of (x); see equation 14
$Z(x)$ = a function of (x); see equation 14
$F_1(k_3)$ = a function of (k_3); see equation 23
U_b = velocity of travel, cm/sec
M = mass rate, gms/sec
P_T = thermal power, $\frac{\text{cal}}{\text{sec}}$
P_1 = surface power density, $\frac{\text{cal}}{\text{sec}}/\text{cm}^2$
$G(P)$ = function of power input during heating; see equation 36
N = number of billets per hour from a single coil with constant power input
N' = same with nonuniform heating
t_3 = total heating time with constant power input, sec
t_3+t_5 = same with nonuniform heating, sec
l_c = length of coil for progressive heating with constant power input, cm
l_c' = same with nonuniform heating, cm

Problems

Problem No. 1. A long cylinder, initially at zero temperature, is supplied with constant power density at the surface, starting at time $t = 0$. It is desired to determine the temperature of the cylinder as a function of radius r for any time t greater than zero. The specification of a long cylinder (in mathematical language an infinite cylinder) indicates that end effects are neglected so that v is a function only of r and t.

Carslaw and Jaeger in their work[1] give the solution of this classical problem. Replacing the constant outward heat flux $-P_0$ in their equation by the constant inward heat flux or power density P_0 gives,

$$v=\frac{P_0 a}{k_t}\left\{\frac{2kt}{a^2}+\frac{r^2}{2a^2}-\frac{1}{4}-2\sum_1^\infty e^{\frac{-k\beta_n^2 t}{a^2}}\times\frac{J_0\left(\beta_n\frac{r}{a}\right)}{\beta_n^2 J_0(\beta_n)}\right\} \quad (1)$$

where $J_0(x)$ is Bessel's function of the first kind and zero order and the β_n's are the positive roots of $J_1(\beta) = 0$. The roots to be substituted in the summation are $\beta_n = 3.83, 7.02, 10.17, 13.2$, etc. It should be noted that this equation is written in terms of the dimensionless variables (r/a) and $T = k\,t/a^2$. This method of writing makes it obvious that one can present the results of the solution in a single set of dimensionless curves as shown in Fig. 1. The abscissa of these curves is the dimensionless time variable T and the ordinates are the quantity inside the parentheses, which is also dimensionless. These may be thought of as dimensionless temperature curves because when the ordinates are multiplied by the quantity $P_0\,a/k_c$ they become actual temperature curves. The same information is plotted in the curves of Fig. 2 to show the temperature distribution in the cylinder for times $T = 0.0125, 0.025, 0.05, 0.10, 0.25$, and 0.50.

The summation term, equation 1, is practically zero for times greater than $T = 0.25$ and the equation becomes

$$v=\frac{P_0 a}{k_c}\left\{2T+\frac{r^2}{2a^2}-\frac{1}{4}\right\} \quad (2)$$

Thus from the time $T = 0.25$ to the end of heating, all parts of the cylinder are increasing in temperature at the same constant rate and the radial temperature distribution is parabolic. This is a true equilibrium condition even though the temperature of the cylinder as a whole continues to increase at a constant rate.

One may substitute $r/a = 1$ and $r/a = 0$ in equation 2 to determine the surface and center temperatures, and the temperature difference $v_s - v_c$ for any value of T greater than 0.25.

$$v_s=\frac{P_0 a}{k_c}\left(2T+\frac{1}{4}\right) \quad (3)$$

$$v_c=\frac{P_0 a}{k_c}\left(2T-\frac{1}{4}\right) \quad (4)$$

$$v_s-v_c=\frac{P_0 a}{2k_c} \quad (5)$$

Equation 2 will be rewritten for future reference in the form,

$$v=v_c+\left(\frac{v_s-v_c}{a^2}\right)r^2 \quad (6)$$

Problem No. 2. Assume heating of the cylinder of problem no. 1 is terminated any time after $T = 0.25$ and it is allowed to equalize with zero heat loss at the surface. Its temperature at time $t = 0$ is given by equation 6 and it is desired to determine temperature as a function of r and t for any time t greater than zero.

The solution of this problem with the initial distribution at time $t = 0$ given as $v = f(r)$ is indicated by Carslaw and Jaeger[2] as follows:

$$v=\frac{2}{a^2}\left\{\int_0^a r'f(r')dr'+\sum_1^\infty e^{-k\alpha_n^2 t}\times\frac{J_0(r\alpha_n)}{J_0^2(a\alpha_n)}\int_0^a r'f(r')J_0(\alpha_n r')dr'\right\} \quad (7)$$

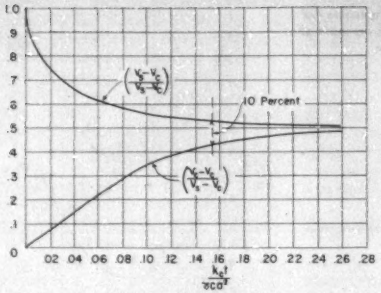


Fig. 3 (left). Temperature equalization in cylinder: power removed

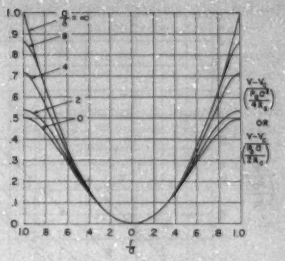


Fig. 4 (right). Temperature distribution in induction-heated billet

where the α_n's are the positive roots of $J_o'(\alpha a) = J_1(\alpha a) = 0$.

The arbitrary function $f(r)$ corresponding to the initial conditions of the present problem is, from equation 6,

$$f(r) = V_c + \left(\frac{V_s - V_c}{a^2}\right) r^2 \tag{8}$$

Capital V's are introduced to distinguish between the initial values of surface and center temperature and the variables v_s and v_c which will be used to represent the corresponding temperatures during equalization. Introducing this $f(r)$ and evaluating the definite integrals gives,

$$v = V_c + (V_s - V_c)\left\{\frac{1}{2} + 4\sum_1^\infty e^{-k\alpha_n^2 t} \times \frac{J_o(\alpha_n r)}{\alpha_n^2 a^2 J_o(\alpha_n a)}\right\} \tag{9}$$

where the α_n's are the positive roots of $J_o'(\alpha a) = J_1(\alpha a) = 0$.

To introduce the same dimensionless variables in this equation as were used in equation 1 write $\beta_n = \alpha_n a$. It follows that $\alpha_n = \beta_n/a$, and $\alpha_n^2 = \beta_n^2/a^2$, and the equation becomes,

$$v = V_c + (V_s - V_c)\left\{\frac{1}{2} + 4\sum_1^\infty e^{\frac{-k\beta_n^2 t}{a^2}} \times \frac{J_o\left(\beta_n \frac{r}{a}\right)}{\beta_n^2 J_o(\beta_n)}\right\} \tag{10}$$

where the β_n's are now the positive roots of $J_o'(\beta) = J_1(\beta) = 0$. These roots are the same as those substituted in equation 1.

Substituting $r = a$ and $r = o$ in equation 10 one obtains equations for surface and center temperature.

$$v_s = V_c + (V_s - V_c)\left\{\frac{1}{2} + 4\sum_1^\infty e^{\frac{-k\beta_n^2 t}{a^2}} \frac{1}{\beta_n^2}\right\} \tag{11}$$

$$v_c = V_c + (V_s - V_c)\left\{\frac{1}{2} + 4\sum_1^\infty e^{\frac{-k\beta_n^2 t}{a^2}} \times \frac{1}{\beta_n^2 J_o(\beta_n)}\right\} \tag{12}$$

The dimensionless quantities inside the parentheses were evaluated and are plotted against the dimensionless time variable $T = k_c t/\gamma c a^2$ in Fig. 3. Actual temperatures are obtained by multiplying the ordinates of the curves by $(V_s - V_c)$ and adding the initial center temperature (V_c).

The Induction-Heated Billet: An Extension of Problem No. 1. Problem no. 1 is quite representative of induction heating when the frequency is high enough to result in a depth of current penetration and a depth of heating small in comparison to billet diameter. It is not representative when depth of current penetration is a considerable fraction of the diameter. The general solution for the temperature distribution in an induction-heated billet would be quite involved and was not found in the literature. The author was able, however, to obtain an accurate restricted solution which is quite satisfactory for the present analysis.

Note that in the solution to problem no. 1 a kind of equilibrium is established after time (t_1) when all parts of the cylinder are increasing in temperature at the same constant rate. This time (t_1) is calculated by

$$T = \frac{k_c t_1}{\gamma c a^2} = 0.25$$

or

$$t_1 = \frac{\gamma c a^2}{4 k_c} \tag{13}$$

It is reasonable to assume that this kind of equilibrium will also occur in an induction-heated billet after a time approximately equal to (t_1). Thus it is permissible to solve the problem of the induction-heated billet on the assumption that all parts of the billet are increasing in temperature at the same constant rate provided the solution is used only for times greater than t_1 as given by equation 13. This limitation is of no consequence in the present analysis because in heating for forging one is primarily interested in temperature distribution toward the end of the heating cycle and because t_1 will always be much less than the total heating time. The problem of the induction-heated billet is discussed below as problem no. 3.

Problem No. 3. Given a long right circular conducting cylinder of radius a heated with constant power input by induction and with negligible heat loss at the surface. Desired to determine the temperature distribution inside the cylinder after sufficient time has elapsed for all parts of the cylinder to be increasing in temperature at the same constant rate. For those interested in the solution of this problem, it is given in the appendix. Only the final equations will be evaluated and discussed in the following.

The general equation for temperature inside the cylinder is

$$v - v_c = \frac{P_a a^2}{4k_c}\left\{\frac{r^2}{a^2} - \frac{1}{k_2}\frac{X\left(k_2 \frac{r}{a}\right) - 1}{Z(k_2)}\right\} \tag{14}$$

where

$$X(x) = \text{ber}^2 x + \text{bei}^2 x$$

$$Z(x) = \text{ber}\, x\ \text{ber}'\, x + \text{bei}\, x\ \text{bei}'\, (x)$$

$$k_2 = k_1 a$$

$$k_1 = \sqrt{\frac{8\pi^2 \mu f 10^{-9}}{\rho}}$$

The quantity inside the parentheses is dimensionless and is expressed in terms of the dimensionless quantities k_2 and r/a. The first is characteristic of the relative depth of current penetration and the second simply expresses per cent

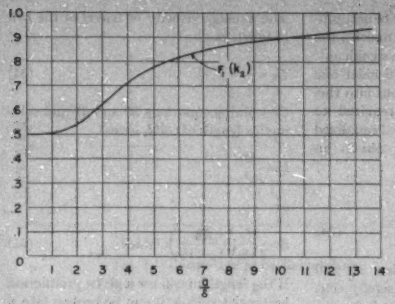


Fig. 5 (left). Correction factor $F_1(k_2)$ for induction-heated billet

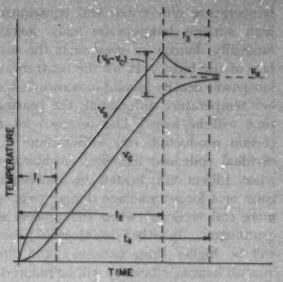


Fig. 6 (right). Typical billet-heating cycle

radius. The accepted definition of depth of current penetration is given by the equation,

$$\delta = \sqrt{\frac{\rho}{4\pi^2 \mu f 10^{-9}}} \qquad (15)$$

It follows therefore that

$$k_2 = k_1 a = \frac{\sqrt{2}a}{\delta} \qquad (16)$$

The value of effective magnetic permeability to be used in the above equations is $\mu = 1$ for so-called nonmagnetic materials. For the so-called magnetic materials like iron and nickel below the Curie temperature it is given by

$$\mu = \frac{1.8 B_s}{H_o} + 1 \qquad (17)$$

where B_s is saturation flux density in gauss and H_o is peak magnetizing force in oersteds. If these metals are heated appreciably above the Curie temperature the value ($\mu = 1$) should be used.

For easy reference and use, the quantity inside the parentheses of equation 14 is evaluated for several values of (a/δ) and plotted in Fig. 4. These may be looked upon as dimensionless temperature curves because the ordinates multiplied by $(P_a a^2/4\ k_c)$ or by $P_o\ a/2\ k_c$ give actual temperature rise above center temperature. Note that the origin of the curves should be thought of as moving upward at a constant rate as all parts of the cylinder continually increase in temperature. Only temperature differences are represented by the curves plotted relative to the stationary axes. It can be shown that the curve $(a/\delta = \infty)$ corresponds exactly to the two upper curves of Fig. 2 derived from problem no. 1 on the basis of constant power input at the surface. Letting (k_2) approach infinity in equation 14 this equation becomes

$$v - v_c = \frac{P_a a^2}{4k_c} \cdot \frac{r^2}{a^2} \qquad (18)$$

If the volume power density P_a is replaced by the equivalent surface power density $P_a = 2\ P_o/a$, this equation becomes

$$v - v_c = \frac{P_o a}{2k_c} \frac{r^2}{a^2} \qquad (19)$$

which is exactly equation 6 of problem no. 1 with $(v_s - v_c)$ replaced by its equivalent $P_o\ a/2\ k_c$ from equation 5.

It is interesting to note how the shape of the temperature curves of Fig. 4 change with the ratio (a/δ). The curve $(a/\delta = \infty$ is one extreme where the depth of current penetration is negligibly small compared to radius and the curve $(a/\delta = 0$ is the other extreme where the depth of current penetration is exceedingly great compared to the radius. Some readers might erroneously conclude that the heating should be nearly uniform in the latter case but of course this is not true because the limiting radial distribution of induced power is parabolic as (a/δ) assumes low values or approaches zero. It is interesting to note that the condition $(a/\delta = 0$ would result in zero heating efficiency. Efficient heating requires (a/δ) to be greater than approximately 1.5. An 8-inch-diameter steel billet induction heated above Curie at a frequency of 60 cycles per second corresponds approximately to $(a/\delta = 1.6$.

In induction heating for forging one is usually satisfied to consider only the temperature differential between surface and center. Substituting $(r/a = 1)$ in equation 14 gives for the surface to center differential,

$$v_s - v_c = \frac{P_a a^2}{4k_c}\left\{1 - \frac{1}{k_2}\ \frac{X(k_2)-1}{Z(k_2)}\right\} \qquad (20)$$

Or upon making the substitution $P_a = 2\ P_o/a$,

$$v_s - v_c = \frac{P_o a}{2k_c}\left\{1 - \frac{1}{k_2}\ \frac{X(k_2)-1}{Z(k_2)}\right\} \qquad (21)$$

Comparing this to equation 5 derived on the basis of constant power input at the surface shows the two equations to be identical except for the quantity inside the parentheses of equation 21 which may be considered as a correction factor which takes into account the fact that heat is generated below the surface. It is convenient to write equation 21 as

$$v_s - v_c = \frac{P_o a}{2k_c} F_1(k_2) \qquad (22)$$

where

$$F_1(k_2) = \left\{1 - \frac{1}{k_2}\ \frac{X(k_2)-1}{Z(k_2)}\right\} \qquad (23)$$

Equation 22 may be used quite generally to calculate the surface to center temperature differential near or at the end of the heating cycle. In fact it may be used to calculate temperature differential for any time greater than t_1 given by equation 13. $F_1\ (k_2)$ is plotted for convenient use in Fig. 5.

Practical Considerations

The aim in induction billet heating is to deliver billets to the press uniformly heated to a certain specified temperature. A radial temperature differential of less than 50 degrees Fahrenheit (F) is usually required. The temperature along the length of a billet must also be reasonably uniform, but this is a function of coil design and/or method of handling and will not be considered in this paper. Billets may be heated individually in individual coils or continuously by intermittently pushing a string of billets through one long coil or a number of coil sections arranged in tandem. If a high power density is used the heating time to temperature will be short but the temperature differential between surface and center at the end of heating will be great. If sufficient heat is supplied to a billet under this condition to allow it to equalize to the desired final temperature, the large

temperature differential and consequent high surface temperature may metallurgically damage or even melt the surface of the billet. If, on the contrary, a low power density is used to accomplish a low temperature differential, the heating time will be long. This requires for a certain production rate either more individual coils and handling mechanisms when billets are heated in individual coils, or a longer machine (longer coils or more coil sections) when the process is continuous. In either case thermal losses will be higher than necessary and the over-all heating efficiency will be reduced. It is clearly desirable therefore to use the highest possible power density consistent with a reasonable maximum surface temperature of billets during heating. The desirable temperature differential dictated by this consideration is usually greater than the low temperature differential required at the press. It is customary therefore to allow some dwell time after power is removed or after a billet emerges from the heating coil or coils for temperature equalization. This may in some cases be simply the time required to transport the billet from the heater to the press. The billets may be large or small, ferrous or nonferrous, and the temperature to which they are heated depends on the metal and subsequent processing. In special cases the maximum power input and temperature differential may be limited by mechanical stresses set up within the billet during heating.

Analysis of Heating Cycle

A typical heating cycle based on constant power input during heating is shown in Fig. 6. At the start of heating, surface temperature increases much more rapidly than center temperature but after a time approximately equal to t_1 (equation 13), the center catches up and thereafter increases at the same rate as the surface temperature. This is in line with the solutions of problem nos. 1 and 3. The temperature of all parts of the billet continue to rise at a constant rate until time t_2, when power is turned off or the billet is removed from the heating zone. The temperature differential at the end of heating is $V_s - V_c$ and the billet equalizes to some low differential in the additional time, t_3. It will be assumed in this paper that the billet has equalized when the differential has decreased to 10% of its value at the end of heating. The mean temperature of the equalized billet is V_a. The total cycle time including equalization but excluding handling time is t_t. Since equalization will usually be accomplished outside the heating coil or coils, the actual heating time t_2 is more significant.

Considering a unit length of billet one may equate the total energy put into this length to the total energy required to raise its temperature to the final equalized temperature V_a and thereby obtain an expression for (t_2).

$$2\pi a P_o t_2 = \pi a^2 \gamma c V_a$$

$$t_2 = \frac{\gamma c a V_a}{2P_o} \tag{24}$$

Substituting P_o from equation 22 and writing $v_s - v_c$ as $V_s - V_c$ to signify that that the temperature difference is specifically that at the end of heating gives,

$$t_2 = \frac{\gamma c a^2}{4k_c}\left(\frac{V_a}{V_s - V_c}\right)F_1(k_2) \tag{25}$$

The equalizing time (t_3) is based on reducing the differential to 10% of its value $V_s - V_c$ at the end of heating. The curves of Fig. 3 show that this condition is met when

$$\frac{k_c t_3}{\gamma c a^2} = 0.154$$

if the temperature distribution prior to equalization is parabolic. It will be assumed that this relationship also predicts with sufficient accuracy the equalizing time from the actual distribution somewhat altered as it is by heat generation below the surface. Therefore,

$$t_3 = \frac{0.154\gamma c a^2}{k_c} \tag{26}$$

It is instructive to express t_1 and t_3 in terms of the heating time t_2. Thus,

$$\frac{t_1}{t_2} = \left(\frac{V_s - V_c}{V_a}\right)\frac{1}{F_1(k_2)} \tag{27}$$

$$\frac{t_3}{t_2} = 0.616\left(\frac{V_s - V_c}{V_a}\right)\frac{1}{F_1(k_2)} \tag{28}$$

It is clear from equations 27 and 28 that with the ratio of differential to final equalized temperature usually required in practical induction heating, both t_1 and t_3 will be small in comparison to heating time t_2.

The number of billets which can be heated per hour in a single coil (one billet in the coil) with equalization outside the coil is

$$N = \frac{3{,}600}{t_2}$$

$$N = \frac{14{,}400k_c}{\gamma c a^2}\left(\frac{V_s - V_c}{V_a}\right)\frac{1}{F_1(k_2)} \tag{29}$$

In the continuous process the production rate is

$$M = U_b \pi a^2 \gamma \tag{30}$$

The average velocity of travel of the row of billets is

$$U_b = \frac{l_c}{t_2} \tag{31}$$

Combining equations 30 and 31 and solving for length of coil l_c gives

$$l_c = \frac{Mt_2}{\pi a^2 \gamma}$$

or

$$l_c = \frac{Mc}{4\pi k_c}\left(\frac{V_a}{V_s - V_c}\right)(F_1(k_2)) \tag{32}$$

If the length of coil for a given production becomes too long, it can be broken into n coils or rows of coils with $1/n$ of the production passing through each. Each coil or row of coils must of course be supplied with separate pusher. When, as sometimes happens, the coil is only a few billets long, special attention must be given to the relative lengths of billet and coil so that an integral number of billets will fall in the coil in such a way as to be properly heated.

The thermal power required to heat billets at a certain mass rate M to equalize at a temperature V_a is

$$P_T = McV_a \tag{33}$$

where c is specific heat averaged over the temperature range involved. To determine the power which must be supplied to the coil or coils, one must add thermal losses and take into account coil efficiency.

Nonuniform Rate of Heating

Thus far it has been tacitly assumed that a billet is subjected to constant power input throughout the heating cycle. This may or may not be the case. The effect of nonuniform heating rate will be illustrated by considering a two stage heating cycle with constant but different power inputs in the two stages. Referring to Fig. 7, curve A corresponds to constant power input and shows the average billet temperature increasing uniformly to the final value V_a in time t_2. The final temperature differential will be characteristic of the power density P_o. Curve B represents a 2-stage heating cycle wherein an increased power density P_1 is applied for an initial period of time t_a with the power density P_o of curve A applied during the final period t_b. Both curves arrive at the same final average temperature V_a and the final temperature differentials will be practically equal provided t_b is greater than t_1 (equation 13). The total heating time $t_a + t_b$ will be less than the total heating

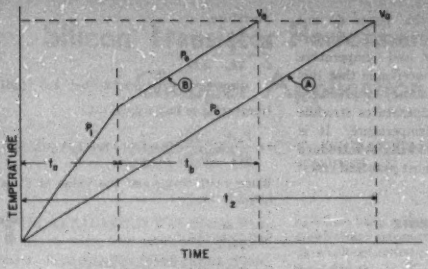


Fig. 7. Comparison of heating cycles

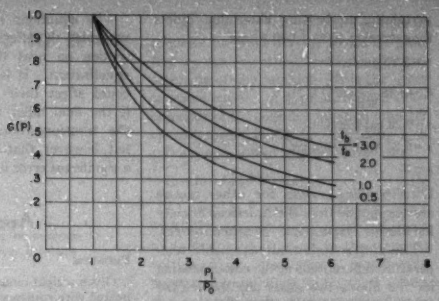


Fig. 8. Correction factor G(P): nonuniform heating

time t_b with constant power input. This is a practical way of reducing total heating time since the higher temperature differential during the early part of the heating cycle is not objectionable and the temperature differential is adjusted to a reasonable value by a reduction in power input well before maximum average billet temperature is reached. The possible shortening of the total heating time by such a 2-stage cycle will now be investigated.

Both heating cycles of Fig. 7 must provide in the billet the same final average temperature V_a, and the power density during the latter part of both cycles must be the same. One may therefore write,

$$P_1 t_a + P_0 t_b = P_0 t_z \tag{34}$$

or

$$t_z = t_b + \frac{P_1}{P_0} t_a \tag{35}$$

Dividing the left and right hand sides of equation 35 into the quantity $t_a + t_b$ gives

$$\frac{t_a + t_b}{t_z} = G(P) = \frac{1 + \frac{t_b}{t_a}}{\frac{P_1}{P_0} + \frac{t_b}{t_a}} \tag{36}$$

It is to be noted that one has free choice in selecting the ratios P_1/P_0 and t_b/t_a and the resulting shortening of the heating time is indicated by equation 36. Substituting t_z from equation 25 the total heating time for the 2-stage heating cycle becomes

$$t_a + t_b = \frac{\gamma c a^2}{4 k_c} \left(\frac{V_a}{V_s - V_c} \right) (F_1(k_2))(G(P)) \tag{37}$$

$G(P)$ is used to represent the right hand side of equation 36. It is evaluated for a number of values of the ratios (P_1/P_0) and (t_b/t_a) and plotted in Fig. 8 for easy reference.

Using the results of the last section one may rewrite equations 29 and 32 to make them applicable for nonuniform heating.

$$N' = \frac{14{,}400 k_c}{\gamma c a^2} \left(\frac{V_s - V_c}{V_a} \right) \left(\frac{1}{F_1(k_2)(G(P))} \right) \tag{38}$$

$$l_s' = \frac{Mc}{4 \pi k_c} \left(\frac{V_a}{V_s - V_c} \right) (F_1(k_2))(G(P)) \tag{39}$$

Equations 37, 38, and 39, calculate respectively the minimum heating time in seconds, the maximum number of billets per hour from a single coil (one billet in the coil), and the minimum length of coil to be used in continuous progressive heating. None of these allow time for handling which must be separately provided. Equalizing time is also to be separately calculated from equation 26.

Discussion of Results

Equations 26, 37, 38, and 39 together with supporting material in the form of dimensionless curves, and the explanation of the derivations, constitute the practical results of this analysis. The author has done sufficient checking of the equations to conclude that they agree quite well in general with practical experience. They have the advantage of being written in terms of the physical properties of the material being heated and consequently are broadly applicable in any case where these properties are known.

Equations are derived on the basis of cylindrical billets but should be approximately applicable to the heating of round cornered square billets if an equivalent diameter is used based on equal areas.

A method of taking into account approximately the effect of thermal loss in depressing surface temperature and reducing temperature differential was worked out but is not included because of lack of space. This effect is usually not important in well-insulated coils with practical heating rates. The effect, however, can become quite large when heating rates are low and surface temperatures are high. So that in extreme cases the surface of the billet may even be cooler than the center at the end of heating.

Equation 38 shows that the number of billets which can be heated per hour in a single coil is directly proportional to the temperature differential which is permitted at the end of heating. The number also depends on the physical properties and size of the billet, on the frequency used, and on the variation of heating rate throughout the heating cycle.

Equation 39 for the length of coil to use in progressive heating shows the same effect of final differential, physical properties, size, etc., but in addition illustrates another interesting fact. It shows that for a given material, heated at a frequency suited to billet diameter, with a given variation of heating throughout the cycle, and with a certain final differential, the length of coil is directly proportional to production rate in gm/sec or lb/hr and is independent of billet diameter. Thus for heating 5,000 lb/hr of steel to forging temperature, the total length of coil will be roughly the same for billets 2 inches in diameter or billets 10 inches in diameter. This is convenient to remember for quick estimating purposes.

Typical values of physical properties to be used in calculations pertaining to the induction heating of steel billets for forging (2,100–2,300 F) are

$\gamma = 7.8$ gm/cm^3

$k_c = 0.07 \frac{\text{cal}}{\text{sec}}$ cm^{-1} C^{-1}

$c = 0.16$ cal/gm/C

$\rho = 120 \times 10^{-6}$ ohm cm

$\mu = 1$

Typical values to be used in calculations

pertaining to heating aluminum billets for extrusion (900–1,000 F) are

$\gamma = 2.76$ gm/cm³

$k_c = 0.38 \frac{\text{cal}}{\text{sec}}$ cm^{-1} C^{-1}

$c = 0.23$ cal/gm/C

$\rho = 9.25 \times 10^{-6}$ ohm cm

$\mu = 1$

Corresponding values for other metals are obtainable from the American Society for Metals "Handbook" or other sources.

The author feels that reasonable temperature differentials at the end of heating for the above two cases are 350 F (194 C) for steel and 150 F (83.5 C) for aluminum. These values are merely illustrative, however, and should be decided each time a new job is approached involving these or any other metal.

Conclusions

Many generalized conclusions could be drawn from the equations but these are so obvious from an inspection of the equations that it seems superfluous to list them. It is important to note that heating time and consequently the number of coils and handling equipments required for a given production rate is critically dependent on the temperature differential permitted at the end of heating. This coupled with the fact that equalizing time is short in comparison to heating time makes it important to choose the differential at the end of heating as great as metallurgical considerations permit.

The equations of this paper permit one completely to plan equipment layout and operational procedure for an induction heating installation for forging or extrusion, including thermal power requirements. Designing the induction heating coils to induce the necessary thermal power is the next step and for this the reader is referred to reference 3 or other sources of information on coil design.

The treatment of this paper is based on classical heat-flow equations and no empirical constants are introduced. The results appear to agree well with practical experience when accurate values of physical properties are substituted in the equations. Accurate values of high temperature physical properties are not easy to find, and the reader is especially warned about values of thermal conductivity which in some alloys increase with temperature instead of decreasing, as is customary in pure metals. Although the introduction of empirical constants has not been found necessary, it may be found later that such would be useful in increasing the accuracy of the equations.

The equations in this paper do not recognize the effect of hests of transformation on heat flow and temperature equalization. It is conceivable that this could affect their accuracy appreciably if the final billet temperatures straddle the transformation temperature. It is felt, however, that this will not introduce a significant error in most practical cases.

Appendix

Problem

Given a right circular cylinder of radius a heated with constant power input by induction. Desired to find the temperature distribution in the cylinder after sufficient time has elapsed for all parts of the cylinder to be increasing in temperature at the same constant rate.

Solution

The conditions of the problem require that the power flow across any internal cylindrical surface of radius r be equal to the power generated inside this radius minus the power required to raise the temperature of the mass of metal inside this radius at the constant rate specified. This is expressed mathematically by

$$-2\pi r k_c \frac{\partial v}{\partial r} = \int_0^r \pi r P dr - \pi r^2 \gamma c \frac{\partial v}{\partial t} \qquad (40)$$

where P is the instantaneous volume power density generated at a point in the cylinder. To evaluate the constant rate of temperature rise in terms of average volume power density P_a,

$$\pi a^2 P_a = \pi a^2 \gamma c \frac{\partial v}{\partial t}$$

and

$$\frac{\partial v}{\partial t} = \frac{P_a}{\gamma c} \qquad (41)$$

Substituting this value in equation 40 gives

$$\frac{dv}{dr} = \frac{P_a r}{2k_c} - \frac{1}{rk_c} \int_0^r r P dr \qquad (42)$$

It can be shown that the power density P in terms of electrical quantities is[4]

$$P = M(\text{ber}'^2\, k_1 r + \text{bei}'^2\, k_1 r) \qquad (43)$$

where

$$k_1 = \sqrt{\frac{8\pi^2 \mu f 10^{-9}}{\rho}}$$

Substituting in equation 42 gives

$$\frac{dv}{dr} = \frac{P_a r}{2k_c} - \frac{M}{rk_c} \int_0^r r(\text{ber}'^2\, k_1 r + \text{bei}'^2\, k_1 r) dr \qquad (44)$$

But,

$$\int_0^r r(\text{ber}'^2\, k_1 r + \text{bei}'^2\, k_1 r) dr = \frac{r}{k_1}(\text{ber}\, k_1 r\ \text{ber}'\, k_1 r + \text{bei}\, k_1 r\ \text{bei}'\, k_1 r) \qquad (45)$$

Therefore,

$$\frac{dv}{dr} = \frac{P_a r}{2k_c} - \frac{M}{k_1 k_c} \times (\text{ber}\, k_1 r\ \text{ber}'\, k_1 r + \text{bei}\, k_1 r\ \text{bei}'\, k_1 r) \qquad (46)$$

Integrating to find v gives,

$$v = \frac{P_a r^2}{4k_c} - \frac{M}{2k_1^2 k_c} (\text{ber}^2\, k_1 r + \text{bei}^2\, k_1 r) + C_1 \qquad (47)$$

When $r = 0$, $v = v_c$, and the value of C_1 is found to be,

$$C_1 = \frac{M}{2k_1^2 k_c} + v_c$$

Substituting in equation 47 gives,

$$v - v_c = \frac{P_a r^2}{4k_c} - \frac{M}{2k_1^2 k_c} \times (\text{ber}^2\, k_1 r + \text{bei}^2\, k_1 r - 1) \qquad (48)$$

It is desirable at this point to find a relationship between P_a and M so that M can be eliminated.

$$P_a = \frac{1}{\pi a^2} \int_0^a 2\pi r P dr \qquad (49)$$

Substituting P from equation 43

$$P_a = \frac{2M}{a^2} \int_0^a r(\text{ber}'^2\, k_1 r + \text{bei}'^2\, k_1 r) dr \qquad (50)$$

$$= \frac{2M}{k_1 a}(\text{ber}\, k_1 a\ \text{ber}'\, k_1 a + \text{bei}\, k_1 a\ \text{bei}'\, k_1 a) \qquad (51)$$

Solving for M, substituting in equation 48, and rearranging gives,

$$v - v_c = \frac{P_a a^2}{4k_c} \left\{ \frac{r^2}{a^2} - \frac{r^2}{a^2} \frac{k_1 a}{k_1^2 r^2} \frac{X(k_1 r) - 1}{Z(k_1 a)} \right\} \qquad (52)$$

where

$$X(x) = \text{ber}^2\, x + \text{bei}^2\, x$$

and

$$Z(x) = \text{ber}\, x\ \text{ber}'\, x + \text{bei}\, x\ \text{bei}'\, x$$

To facilitate the calculation of numerical values and especially to facilitate the plotting of typical dimensionless temperature curves, it is convenient to introduce a new quantity k_2 defined by $k_2 = k_1 a$. Making this substitution equation 52 becomes

$$v - v_c = \frac{P_a a^2}{4k_c} \left\{ \frac{r^2}{a^2} - \frac{1}{k_2} \frac{X\left(k_2 \frac{r}{a}\right) - 1}{Z(k_2)} \right\} \qquad (53)$$

References

1. Conduction of Heat in Solids (book), H. S. Carslaw, J. C. Jaeger. Clarendon Press, Oxford, England, 1947, p. 275.

2. *Ibid.*, p. 178.

3. Design and Calculation of Induction-Heating Coils, R. M. Baker. *AIEE Transactions*, vol. 76, pt. II, Mar. 1957, pp. 31–40.

4. Mathematics Applied to Electrical Engineering (book), A. G. Warren. Chapman and Hall Ltd., London, England, 1941, p. 256.

5. Flow of Heat in Metals (book), J. B. Austin. American Society for Metals, Cleveland, Ohio, 1942.

Silicon Transistor Performance in a Chopper Application

J. GIORGIS
ASSOCIATE MEMBER AIEE

C. C. THOMPSON
NONMEMBER AIEE

THE AVAILABILITY of silicon transistors with low saturation voltages has made possible the development of low-level transistor choppers that are capable of operating over the temperature range of −50 to 150 C (degrees centigrade) with little change in over-all performance. To understand adequately the behavior of silicon transistors in a chopper type of application it is necessary to review the large-signal behavior of junction transistors.

Several papers in the past have been devoted to germanium transistor choppers and to the large-signal theory which explains this mode of operation. The purpose of this paper is to show:

1. That although some of the assumptions used in the analysis of transistor large-signal[1] behavior do not hold true for silicon transistors, the equations derived in this analysis may be used for silicon transistors with good results if some small modifications are made.

2. Methods for measuring or obtaining the pertinent parameters which describe the large signal behavior.

3. The series shunt chopper is generally limited to applications which have a "low" signal impedance.

4. The performance of a low-level series—shunt chopper over an ambient temperature range of −55 to 150 C.

Nomenclature

I_{E0} = Bulk saturation current of emitter junction with zero collector current
I_{C0} = Bulk saturation current of collector junction with zero emitter current
α_N = Grounded base d-c gain in the normal connection
α_I = Grounded base d-c gain in the inverted connection
$p = q/mkT$
q = Charge on electron
m = Constant determined by the injected current density in the base region
k = Boltzmann's constant
T = Absolute temperature, degrees Kelvin
r_C = Collector body resistance
r_E = Emitter body resistance
r_B = Base body resistance
ϕ_C = Voltage across collector junction
ϕ_E = Voltage across emitter junction
$\phi_{BC} = \phi_E - \phi_C$
E_B = Base drive voltage
R_S = Signal source resistance
R_L = Load resistance

Paper 58-88, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 31, 1957; made available for printing November 12, 1957.

J. Giorgis and C. C. Thompson are with the General Electric Company, Schenectady, N. Y.

Transistor Large-Signal Behavior

Practical Considerations

The large-signal behavior of junction transistors is described by Ebers and Moll in thier classic paper[1] by the following three equations:

$$I_E = \frac{I_{E0}}{1-\alpha_N\alpha_I}(e^{p\phi_E}-1) - \frac{\alpha_I I_{C0}}{1-\alpha_N\alpha_I}(e^{p\phi_C}-1) \quad (1)$$

$$I_C = \frac{\alpha_N I_{E0}}{1-\alpha_N\alpha_I}(e^{p\phi_E}-1) + \frac{I_{C0}}{1-\alpha_N\alpha_I}(e^{p\phi_C}-1) \quad (2)$$

$$\alpha_I I_{C0} = \alpha_N I_{E0} \quad (3)$$

where $p = \frac{q}{mkT}$

The polarities of I_{E0} and I_{C0} in equations 1 and 2 are for the direction of current flow in a p-n-p transistor as shown in Fig. 1. For a n-p-n transistor the signs in equations 1 and 2 would be reversed.

Two of the assumptions made in the development of the preceding equations that do not hold for many of the commercially available germanium and silicon transistors are that 1. the resistivities of the semiconductor regions are low, and 2. the injected current densities are low. Some of the primary effects of the higher resistivities and current densities are as follows:

1. The value of m in $p = \frac{q}{mkT}$ is changed from a value of one to a value between one and two.[3]

2. The values of α_N and α_I at a particular current may be different for some conditions of transistor operation than those at which the alphas are measured.

3. Appreciable IR (current resistance) drops are produced by r_B, r_E, and r_C; the emitter, collector, and base body resistances.

The value of p for either the emitter or collector junction may be found by plotting on semilog paper the forward characteristic of the junction (with the other terminal open). A typical forward characteristic of the collector junction (emitter open) is shown in Fig. 2. Equatons 1, 2, and 3 can be solved to give for this case:

$$-I_B = I_C = I_{C0}(e^{p\phi_C}-1) \quad (4)$$

At very low currents, the IR drops within the transistor are small compared to ϕ_C so that ϕ_C is equal to the terminal voltage, V_{BC}. Also, since p is generally between 19 and 40 at 25 C, a few tenths of a volt will make the exponential term much larger than one, so that:

$$I_C \cong I_{C0}e^{p\phi_C} \quad (5)$$

Thus, the V_{BC} versus I_C curve (on semilog paper) should be a straight line over several decades of current, as shown in Fig. 2. The deviation of the curve from a straight line at higher currents is due to the internal IR drops becoming important. Under these conditions, ϕ_C is no longer equal to V_{BC} but is equal to $V_{BC} - I_C(r_B + r_C)$.

The straight portion of the V_{BC} versus I_C curve can be extended to the abscissa. This extrapolation provides a means of finding the bulk I_{C0}. The value of I_{C0} thus obtained agrees very well with the reverse voltage measurement of germanium transistors. This method of measurement gives a better indication of I_{C0} for silicon transistors than the reverse voltage method. The reason is that surface effect can still cloud the reverse voltage measurement even when the surface resistive component is taken into account.

The value of p can be calculated from the linear portion of the curve in Fig. 2 by the equation:

$$p = \frac{\ln\left(\frac{I}{I_{C0}}\right)}{V} \quad (6)$$

It should be pointed out that although it may be possible to draw more than one

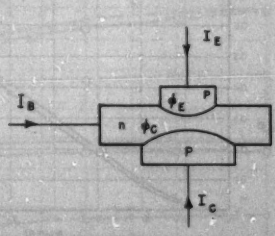


Fig. 1. Generalized geometry of junction transistor

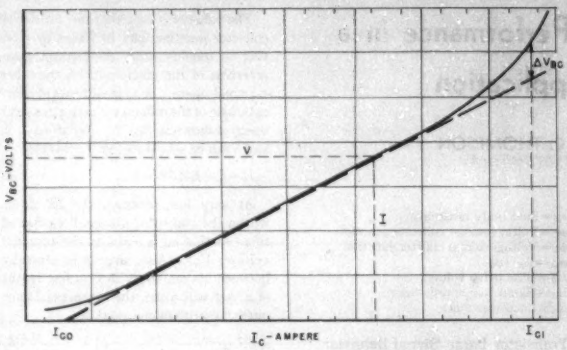


Fig. 2. General forward characteristic of base-collector junction

straight line for some individual silicon transistors, the value of β obtained from each straight line is quite close. However, the value of I_{C0} obtained from each of the straight lines can differ much more because of the exponential nature of equation 4.

The values of the base and collector body resistances (r_C and r_B) can be determined from the V_{BC} versus I_C curve by:

$$r_B + r_C = \frac{\Delta V_{BC}}{I_{C1}} \tag{7}$$

r_C can be measured as outlined in the Appendix and subtracted from equation 7 to give r_B.

In a like manner, by measuring and plotting the emitter junction characteristics (with the collector open), I_{E0} and r_E can also be determined. With I_{E0} and I_{C0} determined, and with α_I and α_N measured (as outlined in the Appendix), equation 3 can then be checked.

Since I_{E0} and I_{C0} can be determined more accurately with germanium transistors than with silicon transistors, better agreement with equation 3 has been obtained with germanium than with silicon transistors.

Figs. 3(A), 3(B), and 3(C) show the forward characteristics of the collector junction (emitter open) for the following types of silicon transistors:

1. Philco surface alloy, *L-6101*, p-n-p.
2. Raytheon alloy, *CK-791*, p-n-p.
3. General Electric meltback, *ZJ-12*, n-p-n.

Table I presents the values, of β, r_C, and I_{C0} for two units of each transistor-type derived by the methods discussed previously. r_B was also determined; typical values were 100 ohms for the *L-6101*, 1,000 ohms for the *CK-791*, and less than 5 ohms for the *ZJ-12*.

Static Properties of a Transistor Switch

In a chopper, the transistor is used as a switch. The emitter and collector termi-

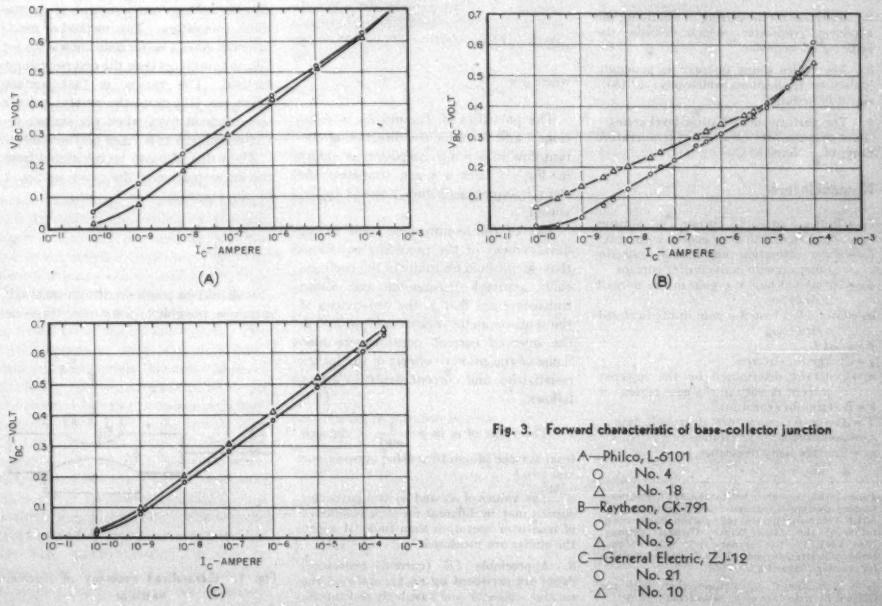


Fig. 3. Forward characteristic of base-collector junction

A—Philco, L-6101
○ No. 4
△ No. 18
B—Raytheon, CK-791
○ No. 6
△ No. 9
C—General Electric, ZJ-12
○ No. 21
△ No. 10

Table I

	L-6101		ZJ12		CK791	
	No. 4	No. 18	No. 10	No. 21	No. 6	No. 9
I_{CO} (max)	21	170	120	110	340	10
P, volts	24.5	21.8	22.5	23	25.7	31.5
r_C, ohms	1.12	0.9	15.75	22.5	0.56	0.3
$\alpha_N (I_C/I_E)$	See Fig. 7					

nals are used as the switch terminals, while either the emitter-base or collector-base terminals is used for opening the switch.

Because the transistor is not a perfect switch in either the open or closed condition, the nature and magnitude of the imperfections must be known as they ultimately determine the over-all performance of the chopper.

Fig. 4 shows a schematic representation of a transistor used as a switch. The transistor switch (emitter-to-collector) is open when the base is made positive with respect to the collector for a p-n-p transistor (or negative for an n-p-n transistor). When the polarity of the voltage between base and collector is reversed, the switch is closed. Notice that the collector terminal is common to both the switching voltage and the load. This arrangement is called the inverted connection and has been shown by Bright[3] to be the best configuration for the low-level chopper application.

Fig. 5 shows the equivalent circuits of the transistor configuration in Fig. 4 for both the open and closed condition. In the open condition, g_{sh} is the "leakage admittance" of the emitter junction, and E_B is the reverse voltage across the collector-base terminal. As long as V_{EC} of Fig. 4 (which is made up of $I_B R_L$ and any external voltage) is less than E_B, the emitter junction will have a reverse voltage across it. Thus $\epsilon^{\phi_E/kT}$ and $\epsilon^{\phi_C/kT} \ll 1$, so that equation 1 may be solved for I_E:

$$-I_E = \left(\frac{\alpha_I}{\alpha_N}\right)\left(\frac{1-\alpha_N}{1-\alpha_I\alpha_N}\right) I_{C0} \qquad (8)$$

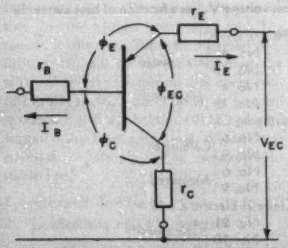


Fig. 4. Diagram of inverted connection

Since $\alpha_I < \alpha_N$, I_E is a constant current which is less than I_{CO}; thus, I_E is shown as a constant-current generator in Fig. 5 for the open condition. At high temperature, the I_{C0} component overshadows the "leakage admittance" component and must be accounted for in chopper design. Fig. 6 shows typical I_{C0} curves as a function of reverse voltage for the three types of silicon transistors at 100 and 150 C.

For the closed equivalent circuit in Fig. 5, r_E and r_C are the emitter and collector body resistances. The battery $I_B r_C$ is a voltage drop caused by the base-collector forward bias current and the collector body resistance. ϕ_{EC} is derived from equations 1, 2, and 3:

$$\phi_{EC} = \frac{1}{p} \ln \left[\frac{1 - \frac{I_E}{I_B}\left(\frac{1-\alpha_I}{\alpha_I}\right)}{(\alpha_N)\left(1 + \frac{I_E}{I_B}(1-\alpha_N)\right)} \right] \qquad (9)$$

In low-level chopper applications where $I_E/I_B \ll 1$, equation 9 reduces to:

$$\phi_{EC} \cong \frac{1}{p} \ln \alpha_N \qquad (10)$$

Equation 10 can be simplified for values of $0.9 < \alpha_N < 1$ by expanding the $\ln \alpha_N$ into a series and neglecting the higher order terms:

$$\phi_{EC} \cong \frac{mkT}{q}(1-\alpha_N) \qquad (11)$$

ϕ_{EC} can be minimized by making α_N as high as possible.

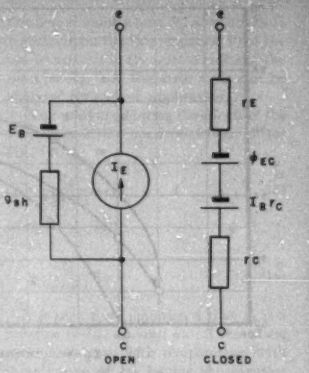


Fig. 5. Equivalent circuit for open and closed conditions

The voltage across the emitter-collector terminals for the closed conditions is:

$$V_{EC} \cong \frac{mkT}{q} \ln \alpha_N + I_B r_C \qquad (12)$$

Since α_N varies with current (see Fig. 7) there is generally one value of base drive where V_{EC} is a minimum. Figs. 8(A), 8(B), and 8(C) show respectively the measured and calculated values of V_{EC} versus I_C for the *L-6101*, *CK-791*, and *ZJ-12* transistors with open emitters. The constants used in the calculations of equation 12 are shown for each transistor in Table I. The values of α_N as used in the calculation were obtained from Fig. 7. Good correlation between calculated and measured points was obtained for currents larger than 0.5 milliampere. At lower currents, the calculated points for the *L-6101* and *ZJ-12* were usually higher than the measure points while the con-

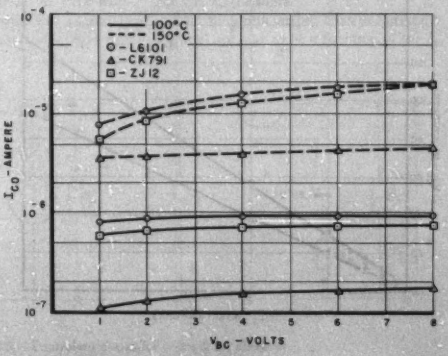


Fig. 6 (right). I_{C0} versus V_C for 100 C and 150 C ambient temperature

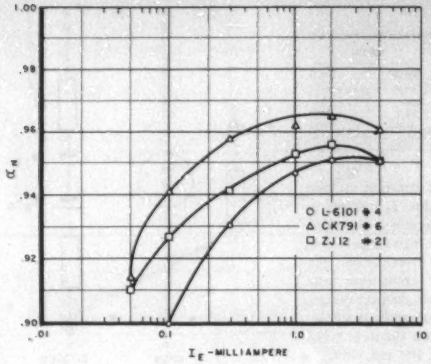
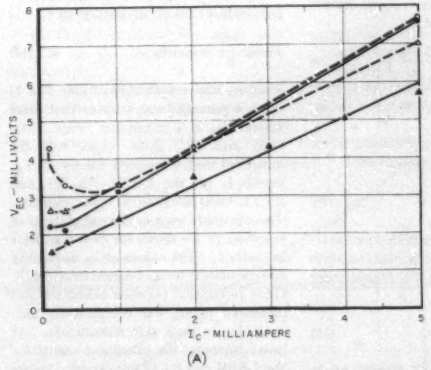


Fig. 7. Normal alpha α_N as a function of collector current I_C

verse was true for the *CK-791*. It might be pointed out that similar effects at very low currents (less than 0.05 milliampere) have been experienced with two different types of germanium transistors which had values of p about 33.

One possible reason for the discrepancy between the calculated and measured values of ϕ_{EC} at low current is that the α_N that exists in the transistor during the ϕ_{EC} measurement is higher than the value used in the calculation. The explanation for this is that during the α_N measurement, the collector and base are essentially short circuited ($V_{CB}=0$) so that only $1-\alpha_N/\alpha_N$ of the collector current flows through the base terminal. However, during the ϕ_{EC} measurement, I_B is zero so that all of I_C flows through the base terminal. The fields thus set up in the base region are then different for the ϕ_{EC} measurement and the alpha measurement.



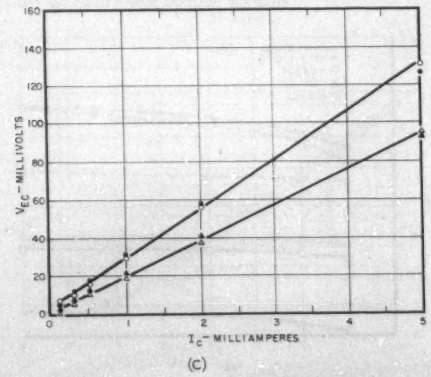


Fig. 8. Emitter-collector voltage V_{EC} as a function of base current I_B

A—Philco, L-6101
- ○ No. 4 } Calculated
- △ No. 18 }
- ● No. 4 } Measured
- ▲ No. 18 }

B—Raytheon, CK-791
- ○ No. 6 } Calculated
- △ No. 9 }
- ● No. 6 } Measured
- ▲ No. 9 }

C—General Electric ZJ-12
- ○ No. 21 } Calculated
- △ No. 10 }
- ● No. 21 } Measured
- ▲ No. 10 }

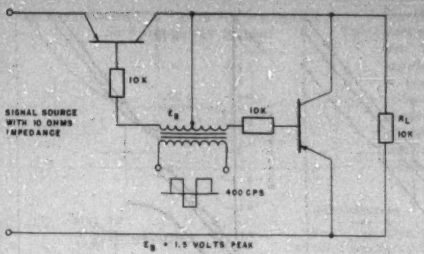


Fig. 9 Circuit diagram of series-shunt chopper

The fields which are established can change the effective emitter area and the current density in the base region, and further, provide additional internal drops which add to or subtract from the ϕ_{EC} voltage. It is possible that the larger field in the base during the ϕ_{EC} measurement increases α_N by changing the surface velocity recombination term of α_N,[3,4] since: 1. The errors are greatest at low currents for both germanium and silicon transistors, and 2. The measured α_N's increase rapidly with the collector current; see Fig. 7.

Evaluation of Chopper

A few transistor chopper circuits are available in the literature.[2,4,5] One of these circuits, referred to as the series-shunt transistor chopper (Fig. 9), was selected for evaluation with silicon transistors in low-level applications.

Although this circuit utilizes two transistors, only the series transistor does the chopping. The shunt transistor is used to minimize and compensate for effects of the series transistor. Each transistor switch is alternately opened and closed at the frequency of the base driving voltage E_S (400 cycles per second). The equivalent circuit for each half-cycle of operation is shown in Figs. 10(A) and (B), respectively.

Chopper performance can be affected in two ways: One is the state of the transistors in the closed condition, and the second is the transistors in the open condition. It will be shown later that the state of the transistors in the open condition will restrict the chopper to low impedance signal applications. In order to simplify the treatment, each condition is covered separately. The assumptions made are:

1. $r_C+r_E<<R_S+R_L$
2. ξ_{sh} shown in Fig. 5 is neglected
3. I_{E1} and I_{E2} are zero when the effects due to transistors in the closed condition are discussed
4. V_{EC1} and V_{EC2} are zero when the effects due transistors in the open condition are discussed.
5. The input signal is zero.

The equations showing the effect of the transistors in the closed conditions are for Fig. 10(A).

$$V_{LA}=\frac{R_L}{R_S+R_L}V_{EC1} \tag{13}$$

and for Fig. 10(B)

$$V_{LB}=V_{EC2} \tag{14}$$

If it is assumed that the voltages V_{EC1} and V_{EC2} are equal, the two output voltages V_{LA} and V_{LB} which occur on alternate half-cycle will differ by a factor of $R_L/(R_S+R_L)$. Over a cycle of operation this difference will result in an a-c voltage or unbalanced output. However, for thermocouple signal detection where $R_S<<R_L$ the two voltages V_{LA} and V_{LB} will be approximately the same.

In applications where the signal-source impedance R_S is comparable in magnitude to R_L, it is necessary to modify the chopper circuit to maintain a balanced output. This is accomplished by inserting a series resistance equal to R_S at the point X in Fig. 10. The equivalent circuits are now symmetrical and the voltages V_{LA} and V_{LB} of equations 13 and 14 have the same coefficient $R_L/(R_S+R_L)$.

The actual value which R_S now can have depends upon the I_E's (equation 8) of the two transistors in the open condition, the value of R_L, and the error than can be tolerated for a given application.

The equations showing the effects of the transistors in the open condition are for Fig. 10(A)

$$I_{LA}=\frac{R_S}{R_S+R_L}I_{E1} \tag{15}$$

and for Fig. 10(B)

$$I_{LB}=\frac{R_S}{R_S+R_L}I_{E1} \tag{16}$$

It is seen from equation 8 that I_E is a function of I_{C0}, α_N, and α_I. These three parameters vary with temperature, with the changes of I_{C0} having the greatest effect. Although the I_{C0} of silicon transistors is in the 10^{-11} to 10^{-9} ampere range at 25 C, it increases with temperature so that it is in the 10^{-4} to 10^{-3} ampere range at 150 C. At this elevated temperature a difference in the I_{C0} of two transistors of 1 microampere can result in an unbalanced chopper output of several millivolts depending upon the values of R_S and R_L. The limitation in the value of R_S thus is determined by both the difference between I_{E1} and I_{E2} with increasing temperature and the error that can be tolerated in the application.

The chopper shown in Fig. 9 was operated with a signal source of 10 ohms impedance and a load of 10,000 ohms. The ratio of R_S/R_L(10/10,000) makes the inaccuracy contribution of I_E small compared to that of V_{EC}.

Shown in Figs. 11(A), 11(B), 11(C) are the transfer characteristics for the chopper using three different makes of transistors. The minimum output level is 10 microvolts for the *L-6101* units, 20 microvolts for the *CK-791* units, and 80 microvolts for the *ZJ-12* units.

The minimum output level or harmonic content appears to be a function of the I_C

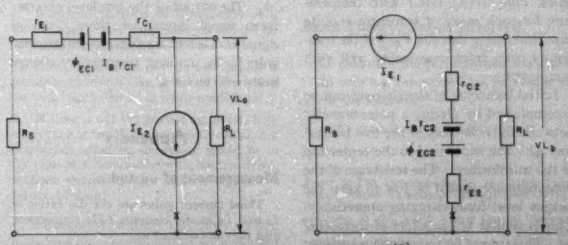
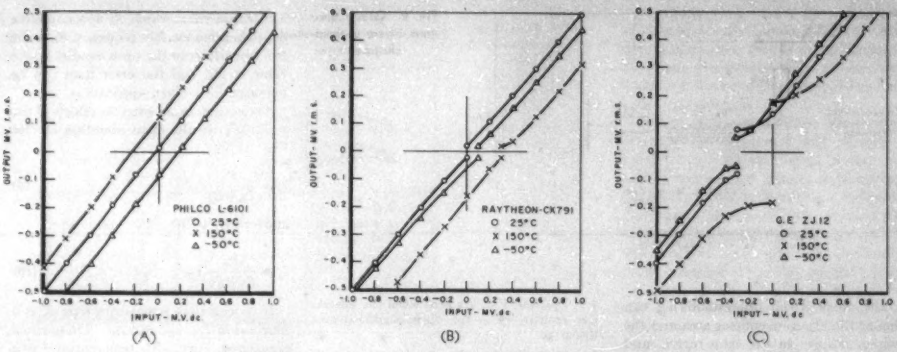


Fig. 10. Equivalent circuit for series-shunt chopper



(A) (B) (C)

Fig. 11. Effect of ambient temperature upon the chopper performance

versus V_{CB} (forward) characteristic and the magnitude of r_C. Considering the I_C versus V_{CB} characteristic, it is important to derive the switching current from a current source since the instantaneous I_C is then independent of the I_C versus V_{CB} characteristic. Because the I_C versus V_{CB} characteristic is temperature dependent, it is important to have a switching-current source that will maintain a constant I_C to minimize the change in V_{EC}. The higher minimum output level noted for the *ZJ-12* transistors is then due to the higher value of r_C and to the difference between the two I_C versus V_{CB} characteristics. The change of level with temperature is a result of the temperature dependence of r_C.

A further consideration in keeping the minimum output level as low as possible is the use of a square-wave current source rather than a sinusoidal one. Since V_{EC} varies with driving current, and since at very low currents this voltage can be large, Figs. 8(A)–8(C), it is important to switch into the optimum operating point in the shortest possible time.

The shift in the input-output characteristics, Figs. 11(A)–11(C) with temperature for each make of transistor is quite small indicating that the V_{EC}'s of the transistors pairs track reasonably well with temperature.

Initial balancing of the chopper can be accomplished by placing a potentiometer between the collectors of the two transistors with the wiper tied to the center-tap of the transformer. The resistance of the potentiometer should be low to keep the output level from changing appreciably. This is similar to the effect of increasing r_C. This type of balance is only necessary when the initial unbalance is out of proportion with the unbalance due to effects of temperature. Matching transistors is another means of obtaining balance if $R_S << R_L$.

Conclusions

1. A small modification of the large-signal equations, and the inclusion of body resistances are necessary to describe the behavior of silicon transistors. Practical techniques are available for determining all the necessary parameters of each transistor.

2. Good correlation between theory and measurement of silicon transistors in the chopper mode of operation was obtained at driving currents greater than 0.5 milliampere for three types of signal transistors tested.

3. The series-shunt transistor chopper is generally limited to "low" signal impedance applications. The maximum value which the signal impedance can have is determined by the magnitude and spread of the I_{CO}'s at the maximum operating temperature and the allowable error for the particular application.

4. The output of the low-level chopper (zero signal input) for three types of signal transistors was less than 500 microvolts for an ambient temperature change from −50 to 150 C.

Appendix

Measurement of α_N and α_I

These current gains are the d-c ratios of I_C and I_E in the common base connection, and they are defined as:

$$\alpha_N = \frac{I_C}{I_E} \text{ when } V_{CB} = 0$$

$$\alpha_I = \frac{I_E}{I_C} \text{ when } V_{EB} = 0$$

Measurement of r_C and r_E

The collector body resistance r_D is determined from equation 12 where:

$$V_{EC} = I_B r_C + \frac{mkT}{q} \ln \alpha_N$$

V_{EC} is measured by a high-impedance voltmeter so that $I_B = 0$. Measurements are taken in the region where $I_B r_C >> \frac{1}{p} \ln \alpha_N$ then:

$$r_C = \frac{\Delta V_{EC}}{\Delta I_B}$$

The emitter body resistance can be determined by reversing the collector and emitter and repeating the preceding measurements.

References

1. Large Signal Behavior of Junction Transistors, **J. J. Ebers, J. L. Moll.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Dec. 1954, pp. 1761–72.

2. Power Rectifiers and Transistors, **R. N. Hall.** *Ibid.*, vol. 40, no. 11, Nov. 1952, pp. 1512–18.

3. Junction Transistors Used as Switches, **R. L. Bright.** *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 111–21.

4. Handbook of Semiconductor Electronics (book), **L. P. Hunter.** McGraw-Hill Book Company, Inc., New York, N. Y., first edition, 1956, pp. 410–11.

5. An Improved High Frequency Transistor, **C. Thorton, R. Roschem, T. Miles.** *Electronic Industries and Tele-Tech*, Philadelphia, Pa., vol. 16, no. 7, July 1957, p. 47.

6. Switching Transistors Used as a Substitute for Mechanical Low-Level Choppers, **A. P. Kruper.** *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 141–49.

7. Transistorized Low-Level Chopper Circuits, **R. B. Hurley.** *Electronic Industries and Tele-Tech*, Dec. 1956, pp. 42–43, 108, 110, 112.

8. Flow Line Analysis, **R. B. Hurley.** *Ibid.*, Apr. 1957, pp. 52–54, 136, 138, 140.

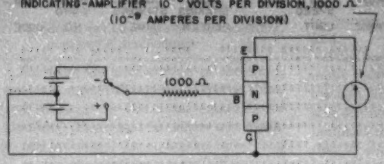


Fig. 12. Circuit for measurement of zero offsets of transistors. Fortunately the meter impedance is high compared to transistor when measuring voltage offset and low comparatively when measuring current offset. For silicon transistors the meter sensitivity was increased to 10^{-13} amperes per division

Discussion

N. E. Polster (Leeds & Northrup Company, Philadelphia, Pa.): In studies reported[1] on transistors used as modulators, we found it useful to call the current injected into the switched circuit under nonconduction conditions "current offset, I_{PI}" and the voltage so injected under conducting conditions "voltage offset, V_{PI}." Fig. 12 shows the simplicity of a circuit for testing for I_{PI} and V_{PI} in the inverted connection. I_{PI} is related to the reverse saturation emitter current I_{EO}, as follows:

$$I_{PI}=\frac{1+\beta_I}{\beta_N}\,I_{EO}$$

where β_I is inverse β
β_N is normal β

Recent measurements made on one of the silicon transistors previously reported[1] give $I_{CO}=1{,}000\times10^{-13}A$ at 25 C and $I_{EO}=I_{PI}=7\times10^{-13}A$. At 65 C, I_{PI} increased to $64\times10^{-13}A$.

I cannot understand the discontinuities in Figs. 11(B) and 11(C). We find agreement with Fig. 11(A) and an important advantage of the transistor modulator is its linearity in the region of zero signal voltage.

Reference

1. Some Advances in Modulators for Precise Measurements, A. J. Williams, Jr., J. U. Eynon, N. E. Polster. *Proceedings*, National Electronic Conference, Chicago, Ill. (To be published)

J. Giorgis and **C. C. Thompson**: The authors wish to thank Mr. Polster for his interest and comments to our paper. The discontinuities of Figs. 11(B) and 11(C) are due to voltage "spikes" which appear across the transistors when they are being switched. Actually, the curves for the Philco *L-6101* transistors, Fig. 11(A), should show a very small discontinuity.

The reason why the *CK-791* transistors produce a larger discontinuity than the *L-6101* transistors is that they have a much lower alpha cut off, thus requiring a larger time to establish a base or collector current at which V_{BC} is low. At low values of collector current V_{BC} is very large, as can be seen from Fig. 8.

The *ZJ-12* on the other hand will switch as fast as the *L-6101*. The discontinuity produced by the *ZJ-12* transistors is due to the large collector body resistance of this type of transistor. During most of the switching time, the magnitude of the base-collector current is determined by the base-collector junction since its impedance is large (at low currents) compared to the 10 kilohms of the source. The $I_{CO}r_C$ drop during this time may be different for the series and shunt transistor and these drops cannot be balanced out.

We would like to point out that the minimum or null voltage (which appears as a discontinuity in Fig. 11) is largely second harmonic. In addition, proper shielding as well as an electrostatic shield between the primary and secondary of the transformer in Fig. 9 is required if the null voltage is to be minimized.

Train Performance and Locomotive Tonnage Ratings Calculated by Digital Computer

J. E. HOGAN
ASSOCIATE MEMBER AIEE

THE FUNCTION of a railroad is to provide transportation. But transportation to be satisfactory and attractive must be not only safe and dependable but must also be accomplished in a reasonable length of time and within the time shown in time tables. To accomplish this, adequate power must be applied to the trains. However, railroads cannot afford to be extravagant. Locomotives are expensive, and the bigger they are the higher the first cost and the cost of operation in wages, fuel, and maintenance.

Some of the typical questions or problems arising in the purchase, selection, and assignment of locomotives and the makeup of trains are as follows:

1. How many cars or how many tons can a certain class of locomotive haul from "A" to "B" in so many hours and minutes?
2. If the time is fixed and the number of cars or tons in a train is specified, what must be the horsepower of the locomotive?
3. If cars are to be added to a train, what will be the effect on the running time?
4. With the addition or elimination of a station stop, what time changes should be made in the schedule?
5. How much time can be cut out of a schedule by eliminating a speed restriction through improvement in track conditions?
6. If time is not the governing factor, how many cars or tons can a locomotive haul at a reasonable speed on the ruling grade, or in the case of a locomotive with electric drive, without damaging the electric apparatus by overheating?

The answers to these and other related questions may be important with respect to potential purchases and utilization of locomotives and other railroad equipment or the investment in track improvements.

Of course, in many cases answers can be obtained through observation, by road tests, that is, by the "trial and error" method. But this may prove expensive to the railroad and possibly annoying to its customers. And, with road tests, when normal operations are most desired, unusual or abnormal conditions may arise to spoil the tests.

Mathematical solutions have been found to be quite satisfactory in most cases. The factors involved are fairly well established, particularly in the case of electric and diesel-electric motive power. The output of such equipment can be predicted reliably.

Basic Theory, Terms Involved

To calculate the performance of a given train over a given route, it is necessary to know the following:

1. The grades, curves, and distances.

Paper 58-100, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted September 5, 1957; made available for printing November 14, 1957.

J. B. Hogan is with the Pennsylvania Railroad Company, Philadelphia, Pa.

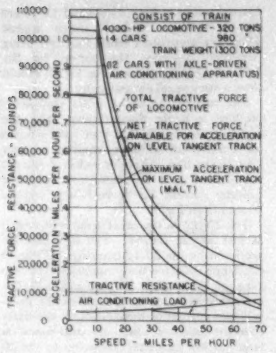


Fig. 1. Typical passenger train data

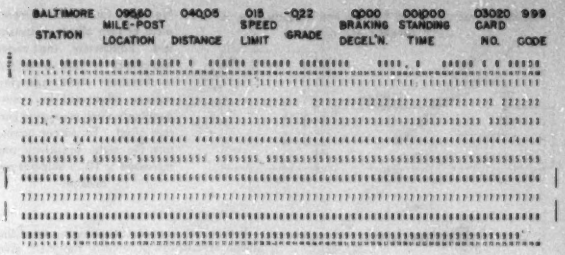


Fig. 2. Track data card

2. The speed restrictions.
3. The weight and tractive resistance of the train.
4. The tractive force of the motive power.
5. The rate of deceleration during braking.
6. The number, location, and duration of station stops.

By *tractive force* is meant the total propelling force at the rims of the driving wheels. The portion of tractive force in excess of that required to overcome tractive resistance is available for producing acceleration and is known as *net tractive force*.

The tractive resistance may be divided into train resistance and track resistance. Several empirical equations for train resistance have been developed, generally of the type:

$$R = K_1 + K_2V + K_3V^2 \qquad (1)$$

where

R = resistance
K_1 = a term representing journal friction
K_2V = a term comprising resistances due to flange friction, swaying and miscellaneous frictions proportional to speed
K_3V^2 = a term giving air or wind resistance varying as the square of the speed

The Davis Formulas for train resistance are in general use in this country.[1] They were developed from tests conducted on several railroads in the 1920's. Resistance calculated by these formulas is somewhat high, however, when applied to the conditions of heavier rail, better road bed, better equipment, and the higher speeds of our railroads today.

Track resistances are the gravitational resistances caused by grades and frictional resistances caused by curves. A train moving up a grade of 1%, or 1 foot rise in 100 feet, must overcome a component of gravity equal to 20 pounds for each 2,000 pounds of train weight. Similarly, when a train moves down a grade of 1%, a component of gravity exerts a force of 20 pounds per ton to assist the movement.

A train moving on a curve is retarded by a force proportional to track curvature. For freight trains this retarding force is usually taken to be 0.8 pound per ton of train weight per degree of curve. It may be considered as the effect of an adverse grade of 0.04% per degree of curvature. Curve resistance is much less for passenger trains and is usually considered negligible.

The maximum tractive force which may be developed by an electric or a diesel-electric motive power unit is determined by the weight on driving wheels and the coefficient of adhesion between the wheels and the rails. Tractive force equivalent to about 25% of the weight on drivers is about the maximum that can be utilized in starting a train on wet-sanded rail.

Adhesion may vary widely and depends on the condition of the rails, whether clean and dry, or wet and perhaps oily. It has been found by tests to vary inversely as the speed. Adhesion factors on wet rail, sanded, according to Candee, indicate that at 11 miles per hour the expected adhesion is about 18%; at 30 miles per hour, 14%.[2] Tractive force in excess of the maximum that the adhesion will sustain will cause the drivers to slip.

The calculations of acceleration are based on Newton's Second Law of Motion which states that the acceleration of a given body is proportional to the force producing it. If the net tractive force is in pounds, the weight W in tons and the corresponding acceleration a is in miles per hour per second, the equation becomes:

$$a = \text{Net Tractive Force}/91.1W \qquad (2)$$

However, this does not allow for the force required to accelerate the rotating parts. Making an allowance for this, the equation in common use is:

$$a = \text{Net Tractive Force}/100\ W \qquad (3)$$

In other words, a net tractive force of 100 pounds per ton will accelerate a train at a rate of 1 mile per hour per second.

As the net tractive force, the force available to produce acceleration, is the force in excess of all the resistance forces that must be overcome, the equation may be written for acceleration on level tangent track as follows:

$$a = \frac{TF - (R_L + R_C + R_A)}{100\ W} \qquad (4)$$

where

a = acceleration on level tangent track in miles per hour per second (negative for retardation)
TF = total tractive force, pounds
R_L = resistance of locomotive, pounds
R_C = resistance of cars, pounds
R_A = resistance caused by axle-driven generators for lighting and air conditioning, pounds
W = total weight of train, tons

For convenience we have abbreviated maximum acceleration on level tangent track to MALT.

Fig. 1 shows, for level tangent track, graphs of total tractive force of a 4,000-horsepower diesel-electric locomotive, total train resistance of the locomotive and 14 passenger-train cars, the resistance caused by 12 of the 14 cars having axle-driven generators supplying power for the lighting and air conditioning load, and the resulting MALT for the train.

As the effect of a grade is to exert a force of 20 pounds per ton of train weight for each per cent of grade, tending either to decrease or to increase the speed of a train, the actual acceleration of a train operating with full throttle is:

$$a = \text{MALT} - 20\ G/100 \qquad (5)$$

where

G = grade, in per cent (positive if ascending, negative if descending)

Knowing the acceleration of the train, we can compute changes in velocity, time, and distance by the usual equations of kinetics. Expressed in terms applicable to our problems, the equations most used in these computations are as follows:

$$\text{Accelerating:} \begin{cases} \Delta V = 60a\ \Delta t & (6) \\ \Delta x = (V + \Delta V/2)\Delta t/60 & (7) \end{cases}$$

$$\text{Constant Speed:} \begin{cases} \Delta t = 60\Delta x/V & (8) \\ \Delta x = V\Delta t/60 & (9) \end{cases}$$

$$\text{Braking:} \begin{cases} \Delta t = (V_1 - V_2)/60b & (10) \\ \Delta x = (V_1^2 - V_2^2)/7{,}200b & (11) \end{cases}$$

where

t = time, minutes
V = speed, miles per hour (mph)
a = acceleration, mph per second
b = braking deceleration, mph per second
x = distance, miles

However, with each change in speed, there is corresponding change in the force producing that change. Therefore, to obtain accurate results, small increments of either velocity, distance, or time must be used and the results calculated step by step. This is a tedious and time-consuming process if done with slide rule and graphs. Each step in the process involves checking for every change in track data such as grades, curves, speed restrictions, etc.

The Pennsylvania Railroad Train Performance Calculator

The Pennsylvania Railroad has an analog computer for calculating train performance.[2] It consists of three curve-drawing instruments, electrically interconnected, and a low-energy auxiliary circuit. The latter contains apparatus necessary to represent by small voltages the effective forces which influence the movement of a train. The MALT data for a given train are incorporated into the circuits of the machine by means of dials on the face of a panel before performance calculations are begun. Upon a chart the machine draws the developed curve of speed versus distance, at the same time producing two other charts showing acceleration versus time, and speed versus time. The operator makes adjustments in the controls as called for by the track data. Results are read from the three charts, which show for every instant of the run, the time, velocity, acceleration, and distance. This method has the advantage of recording the results in graphic form, producing charts by which the train movement over the route can be visualized.

The simplicity of its control, low energy input (300 watts maximum), relatively low cost, and quiet operation make it suitable for use in an engineering office. It performs the work normally in train time, that is, within the time required for the actual train to make the run. Extensive use has been made of this machine since its installation in April 1950. It has been particularly useful during the acquisition of new locomotives and new trains, also for planning and analyzing the movement of modern, high-speed freight trains on fast schedules. Results have been satisfactory and have been checked and found to conform closely with actual operation.

Calculations by Digital Computer

The Pennsylvania Railroad has recently applied a digital computer, the International Business Machines (IBM) model *650* data processing machine, to the calculation of train performance, fuel or electric energy consumption in moving a train, and locomotive tonnage ratings.

The basic "programs" or lists of instructions for the application of the digital computer to these calculations were developed by Dr. E. R. Mullins, Jr. (Swarthmore College and IBM) and Dr. D. L. Thomsen, Jr. (IBM). Input and output are in the form of punched cards. Fig. 2 is a typical track data card.

Calculation of Train Performance

For train performance calculations, three sets of punched input cards are used: 1. the program, about 400 instructions; 2. the train MALT data, calculated beforehand for the train to be run, five cards; and 3. the track data, average about one card per mile of track.

The program of instructions and the train MALT data are loaded on the *650* memory storage drum; the track data cards are read consecutively and automatically by the machine as the calculations proceed.

Fig. 3 shows the information conveyed by 13 typical track data cards. In a typical problem, a train is to be moved from "A" to "B" and the minimum time required to accomplish this is to be calculated. One punched card is used for every point of change in track data and supplies the data ahead of it to the point represented by the next card. Card no. 1.00 says the train leaving that point, mile post location 10.00 is on a grade of 0%, and that its speed must not exceed 30 mph. Card no. 1.10 says the grade beyond mile post location 9.60 is still zero, but that the speed may now increase to 80 mph, and so on.

Most of the track data can be obtained readily from the track charts used by railroads to provide in convenient form information such as this. However, some additional information is needed in connection with braking, that is, reducing speed or stopping.

The rate of deceleration must be specified. It is dependent on such things as grade, weight on axles, wheel diameter, brake shoe contact pressure, area of contact, etc. Fairly accurate values have been determined for passenger trains, varying from about one mile per hour per second for most locomotive hauled trains to about two miles per hour per second for

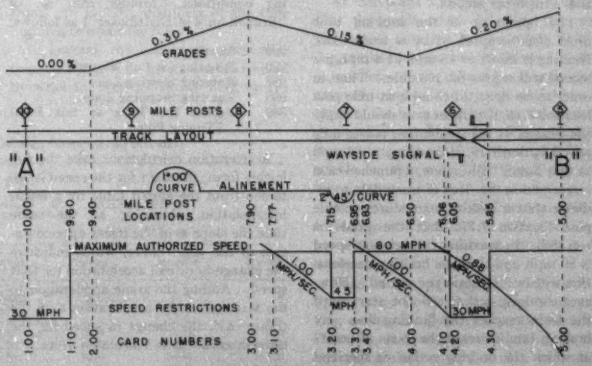


Fig. 3. Essentials of track data

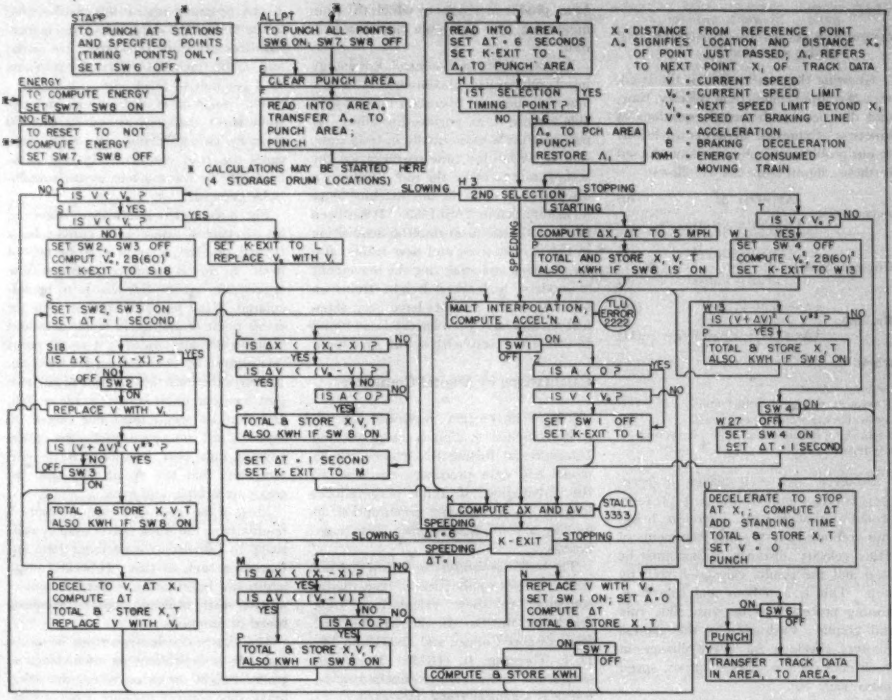

STAPP
TO PUNCH AT STATIONS AND SPECIFIED POINTS (TIMING POINTS) ONLY, SET SW 6 OFF
ALLPT
TO PUNCH ALL POINTS SW6 ON, SW7 SW8 OFF
ENERGY
TO COMPUTE ENERGY SET SW7, SW8 ON
NO-EN
TO RESET TO NOT COMPUTE ENERGY SET SW7, SW8 OFF
CLEAR PUNCH AREA
READ INTO AREA$_0$ TRANSFER Λ_0 TO PUNCH AREA; PUNCH
CALCULATIONS MAY BE STARTED HERE (4 STORAGE DRUM LOCATIONS)
READ INTO AREA; SET Δt = 6 SECONDS SET K-EXIT TO L Λ_1 TO PUNCH AREA
IST SELECTION TIMING POINT ?
YES
NO
Λ_0 TO PCH AREA PUNCH RESTORE Λ_1
X = DISTANCE FROM REFERENCE POINT
Λ_0 SIGNIFIES LOCATION AND DISTANCE X_0 OF POINT JUST PASSED; Λ_1 REFERS TO NEXT POINT X_1 OF TRACK DATA
V = CURRENT SPEED
V_0 = CURRENT SPEED LIMIT
V_1 = NEXT SPEED LIMIT BEYOND X_1
V^* = SPEED AT BRAKING LINE
A = ACCELERATION
B = BRAKING DECELERATION
KWH = ENERGY CONSUMED MOVING TRAIN
IS V < V_0 ?
SLOWING
2ND SELECTION
STOPPING
STARTING
SET SW2, SW3 OFF COMPUT V_0^2, 2B(60)^2 SET K-EXIT TO S18
SET K-EXIT TO L REPLACE V_0 WITH V
SPEEDING
COMPUTE ΔX, ΔT TO 5 MPH
TOTAL AND STORE X, V, T ALSO KWH IF SW8 IS ON
SET SW 4 OFF COMPUTE V_0^2, 2B(60)^2 SET K-EXIT TO W13
SET SW2, SW3 ON SET ΔT = 1 SECOND
MALT INTERPOLATION, COMPUTE ACCEL'N A
TLU ERROR 2222
IS ΔX < (X_1 - X) ?
IS ΔV < (V_0 - V) ?
IS A < 0 ?
SW1
SW2
SW3
SW4
SW6
SW7
REPLACE V WITH V_0
IS (V + ΔV)^2 < V^{*2} ?
TOTAL & STORE X, V, T ALSO KWH IF SW8 ON
SET SW 1 OFF SET K-EXIT TO L
SET ΔT = 1 SECOND SET K-EXIT TO M
IS (V+ΔV)^2 < V^{*2} ?
SET SW 4 ON SET ΔT = 1 SECOND
COMPUTE ΔX AND ΔV
STALL 3333
K-EXIT
DECELERATE TO STOP AT X_1; COMPUTE ΔT ADD STANDING TIME TOTAL & STORE X, T SET V = 0 PUNCH
DECEL TO V_1 AT X_1 COMPUTE ΔT TOTAL & STORE X, T REPLACE V WITH V_1
REPLACE V WITH V_0 SET SW 1 ON; SET A=0 COMPUTE ΔT TOTAL & STORE X, T
PUNCH
TRANSFER TRACK DATA IN AREA, TO AREA$_0$
COMPUTE & STORE KWH

Fig. 4. Flow chart, train performance calculations program

electric multiple-unit trains. Braking deceleration for freight trains varies widely but a conservative value in general use is 1/4 mph per second.

The beginning of the braking zone must also be specified by a track card. Braking from 80 to 45 mph at 1 mph per second will require 62/100 mile. Thus, in order to be down to 45 mph at mile post location 7.15, the engineman should apply the brakes as the locomotive passes mile post location 7.77, if traveling at 80 mph at that point. Therefore, a punched card is inserted in the deck of track data to indicate that a braking zone starts at mile post location 7.77, that the grade is −0.15%, the maximum permissible speed is 80 mph and that the braking deceleration will be 1.00 mph per second. However, braking action will not start until the train reaches the braking line. By braking line is meant the locus of points at which the braking action as specified will cause the train to decelerate to be precisely at the desired speed at the desired location.

The code is used to identify the operating conditions (through the use of "Branch on 8 in Distributer") as follows:

999	Starting
989	Speeding (or Full Throttle)
988	Entering Slowing Zone
998	Entering Stopping Zone
889	Record the Time at this Point (Timing Point)

Acceleration calculations take the following form: the MALT for the speed is obtained from the stored table by a linear interpolation, obtaining at the same time the slope m of the linear approximation of the MALT curve, which indicates the change in rate of acceleration for that speed. Adding the grade acceleration to the MALT gives the acceleration a of the train. ΔV, the change in speed for the time interval Δt is calculated from:

$$\Delta V = a(1+m\Delta t/2)\Delta t \qquad (12)$$

where

ΔV = change in speed, mph
a = acceleration, mph per second
m = slope of linear approximation of MALT curve
Δt = the time increment, seconds

The calculations on the digital computer are carried out by first using a primary value for Δt. At the completion of each calculation of ΔV and Δx, the totals $V+\Delta V$ and $x+\Delta x$ are checked against the track data to ascertain whether the values thus obtained would result in a speed greater than authorized or a distance greater than that of the next change in track data. If neither of these limits will have been exceeded, the totals are recorded and the computation cycle is continued. If either of these limits has been reached, the last calculation is not totaled and a second cycle of computations is begun using a secondary value for Δt considerably less than the primary value. However, when the limit is

reached, using the secondary value for Δt, the cycle is terminated.

If the limiting speed has been attained, time is calculated using that speed and using the distance to the next change in track data. If the limiting distance has been reached, data from the next card are read automatically into the machine and the calculations are resumed using the primary value for Δt.

Whenever the train enters a braking zone, the calculations take a different form. Tests are made to ascertain the time, place and speed at which the train arrives at the braking line with the smaller secondary value of Δt. Braking action is then started, deceleration taking place at the rate b specified.

Even after the train enters the braking zone, operation will continue at full throttle until reaching the braking line. An exception to this occurs if the speed is below that of the speed restriction approached. In this case the speed of the train in the braking zone will not exceed the speed restriction ahead.

Calculation of Consumption of Fuel or Electric Energy

If desired, the energy consumed in moving the train is computed as the train performance calculations progress. To accomplish this, the following additional train information (seven cards) is added to the train data, 1. total weight of the train, 2. a table of total tractive force versus speed, and 3. a table of total train resistance versus speed.

Energy calculations are based on the equation:

$$E = 0.002(TF)\Delta x \qquad (13)$$

where

E = total energy delivered to the rims of the drivers, in kw-hr
TF = total tractive force, in pounds
Δx = distance traversed, in miles

In the case of full-throttle operation, the tractive force is obtained from a table of maximum tractive force versus speed. For partial throttle at constant speed the total resistance is used to indicate the tractive force. During drifting or braking no propelling power is supplied.

The energy thus calculated is the energy delivered at the rims of the driving wheels. By the use of the proper efficiency or conversion factor, the energy input at the pantograph or amount of fuel consumed can be derived.

Fig. 4 shows the flow chart for the calculation of train performance and energy consumption. The program may be entered at several different locations on the memory storage drum, as follows:

ALLPT (punch all points). This will cause an output card to be punched at every location designated by a track card. It is used principally in providing a comparison with results calculated by other methods. By its use a rough graph can be plotted of speed vs. time or distance.

STAPP (punch stops and passing points). This is self-explanatory. It is the method most frequently used.

ENERGY (compute energy). This is used when energy calculations are desired.

NO-EN is used to reset the program to normal, that is, without calculating energy.

The results recorded by holes punched in the output cards are converted into printed tabular form by passing the punched cards through a listing machine. Typical tabulated output data are shown in Table I.

Calculation of Locomotive Tonnage Ratings

The calculations of maximum permissible ratings for locomotives are handled in a somewhat different manner. In these problems we are not concerned with train operating time. We want to know the maximum tonnage a locomotive can haul from "A" to "B" regardless of the time required to make the run.

In general, the maximum permissible tonnage rating for a given route is limited by what the locomotive can haul up the ruling grade. By "ruling grade" is meant the most difficult grade the train will encounter on that route. That sounds like a rather simple definition. However, with the exception of a long uniform grade, it is not always easy to identify the ruling grade and to ascertain its average steepness or its length, particularly in what may be called rolling terrain.

The ruling grade is not always the steepest grade. The steepest grade may be only a few hundred feet in length. Generally speaking, the answer to the question, "What is a ruling grade?" depends on: 1. steepness, 2. length, and 3. conditions of grade and of train speed approaching the foot of the grade in question.

This introduces the subject of "momentum operation." Obviously the faster a train is moving as it reaches the foot of a steep grade, the greater is the kinetic energy stored in the train to assist it in surmounting that grade. To be "ruling," a grade must be long and steep enough to reduce the speed of the train to what we may call the locomotive "rating speed." If the lowest speed on the grade is greater than the rating speed, the full tonnage rating of the locomotive is not being utilized.

The rating speed may be determined by one of several limitations, but it is usually the lowest speed at which the locomotive can be operated for an unlimited time at full throttle. A lower speed with corresponding higher tractive force (and higher motor current in the case of locomotives with electric transmissions) is not practicable, as it would cause slipping of driving wheels or damage to electrical insulation through overheating.

At this point it might be well to note that the grade which is ruling for one locomotive may not be the ruling grade for another locomotive of different characteristics. In the case of the long freight train of today hauled by a powerful locomotive, the train may be longer than the grade which was ruling for a less powerful locomotive. In other words, on a rolling profile, a train may "overhang" one or more grades.

As a corollary to this, the tonnage rating for a two-unit locomotive will often be more than twice that for a single unit, and so on.

Thus, it becomes apparent that establishing accurate tonnage ratings is not an easy matter of applying a single formula. However, it is exceedingly important to the economical operation of a railroad. Tonnage ratings too low will mean un-

Table I. Output Data from Train Performance Calculation

TRAIN: No. 152 LOCOMOTIVE: GG-1
FROM: Washington TO: New York VIA:
TONNAGE: 1,200 CARS: 18 Air-Conditioned: 16 RAILS: Dry
REMARKS: Present Speed Restrictions DATE: 8-26-57

Station	Mile Post	Distance	Speed	Minutes	Kw-Hr
Washington	135.85	0.00	0.0	0.0	0
New York Avenue	134.95	0.90	17.1	4.1	65
Landover	128.70	7.15	80.0	10.4	466
Bowie	120.70	15.15	79.9	16.5	820
Odenton	113.50	22.28	80.0	21.9	1,146
Fulton	97.50	38.15	80.0	34.8	1,529
Baltimore	95.60	40.05	0.0	39.6	1,529

necessarily large investments in motive power with related larger maintenance facilities and greater crew and other operating costs. Tonnage ratings too high, if used, will result in more frequent equipment failures and detentions, slow over-all operation, and high maintenance costs as the result of damaged traction motors or generators.

Present "long-hand" methods of calculating tonnage ratings involve the selection of what is apparently the ruling grade from inspection of the track chart. Tonnage rating is then calculated by the equation:

$$T = \frac{P - 20Gw}{r + 20G} \qquad (14)$$

where

T = tonnage, trailing tons gross weight
P = drawbar pull, in pounds
G = grade, in per cent
w = weight of locomotive, in tons
r = tractive resistance of cars, in pounds per ton

Studies made of a variety of typical conditions involving momentum operation over a rolling profile demonstrated that, with rated tonnage, an averaged grade having a length *approximately* twice that of the train would, in general, result in a reduction in speed to about the minimum rating speed. That is to say, with momentum operation, to be ruling, such grade should be at least twice the length of the train. However, this rule is not exact and tonnage ratings thus obtained should be checked by other means.

Obviously, to be ruling, without momentum operation, the averaged grade must be at least as long as the train.

Checking tonnages thus calculated can hardly be accomplished manually by the step-by-step process because of the tremendous number of computations involved. It can be done with the Pennsylvania Railroad Train Performance Calculator.[3] But even with this analog computer the process requires too much time to check many ratings. The rapidity of computations with the digital computer makes possible a method of quickly and easily calculating tonnage ratings accurately.

There is one major difference between calculating train performance and calculating tonnage ratings. In train performance, to simplify the work, the weight of the train is almost always assumed to be concentrated in one point. This assumption introduces certain errors. Calculations on this basis indicate that the train begins to lose appreciable speed at the instant the locomotive reaches the foot of a steep grade. Actually the grade will not have its full effect on reducing the train speed until the entire train is on the grade. However, the reverse effect takes place as the train passes over a summit. Over the length of runs usually considered in train performance calculations these two types of error cancel each other nicely, and sufficiently accurate values of total time are obtained.

In tonnage rating calculations, however, the length of the train becomes an item of prime importance. The train will be

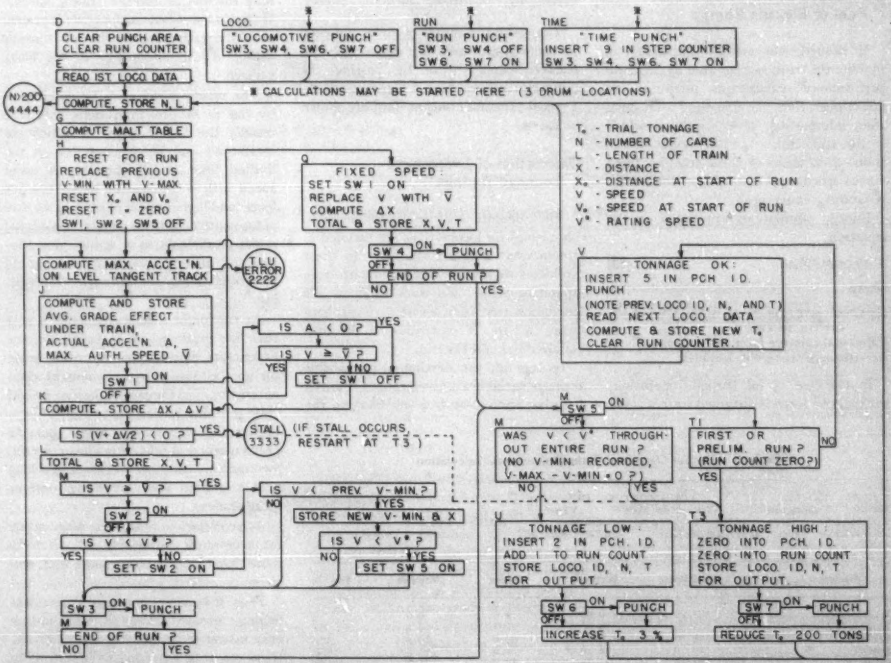

LOCO.
"LOCOMOTIVE PUNCH" SW3, SW4, SW6, SW7 OFF
RUN
"RUN PUNCH" SW3, SW4 OFF SW6, SW7 ON
TIME
"TIME PUNCH" INSERT 9 IN STEP COUNTER SW3, SW4, SW6, SW7 ON
D
CLEAR PUNCH AREA CLEAR RUN COUNTER
E
READ IST LOCO DATA
F
NO 200 4444
COMPUTE, STORE N, L
CALCULATIONS MAY BE STARTED HERE (3 DRUM LOCATIONS)
G
COMPUTE MALT TABLE
H
RESET FOR RUN REPLACE PREVIOUS V-MIN WITH V-MAX RESET X_o AND V_o RESET T = ZERO SW1, SW2, SW5 OFF
Q
FIXED SPEED: SET SW1 ON REPLACE V WITH $\bar{V}$ COMPUTE ΔX TOTAL & STORE X, V, T
T_e - TRIAL TONNAGE
N - NUMBER OF CARS
L - LENGTH OF TRAIN
X - DISTANCE
X_o - DISTANCE AT START OF RUN
V - SPEED
V_o - SPEED AT START OF RUN
V^* - RATING SPEED
I
COMPUTE MAX. ACCEL'N. ON LEVEL TANGENT TRACK
T.L.U. ERROR 2222
SW 4 ON PUNCH
OFF END OF RUN ?
NO YES
V
TONNAGE OK: INSERT 5 IN PCH I.D. PUNCH (NOTE PREV. LOCO I.D., N, AND T) READ NEXT LOCO DATA COMPUTE & STORE NEW T_e CLEAR RUN COUNTER
J
COMPUTE AND STORE AVG. GRADE EFFECT UNDER TRAIN, ACTUAL ACCEL'N A, MAX. AUTH. SPEED $\bar{V}$
IS A < 0 ?
YES
IS V ≥ $\bar{V}$?
YES
NO
SET SW1 OFF
SW 1 ON
OFF
K
COMPUTE, STORE ΔX, ΔV
IS (V + ΔV/2) < 0 ?
YES
STALL 3333
(IF STALL OCCURS RESTART AT T3)
L
TOTAL & STORE X, V, T
M
IS V ≥ $\bar{V}$?
YES
IS V < PREV. V-MIN ?
NO
YES
STORE NEW V-MIN & X
SW 2 ON
OFF
IS V < V^* ?
YES
NO
SET SW2 ON
IS V < V^* ?
NO
YES
SET SW 5 ON
SW 3 ON PUNCH
M
END OF RUN ?
NO
YES
M
SW 5 ON
OFF
M
WAS V < V^* THROUGHOUT ENTIRE RUN ? (NO V-MIN. RECORDED, V-MAX - V-MIN < 0 ?)
NO
YES
T1
FIRST OR PRELIM. RUN ? (RUN COUNT ZERO?)
NO
YES
U
TONNAGE LOW: INSERT 2 IN PCH. I.D. ADD 1 TO RUN COUNT STORE LOCO. ID, N, T FOR OUTPUT
T3
TONNAGE HIGH: ZERO INTO PCH. I.D. ZERO INTO RUN COUNT STORE LOCO. ID, N, T FOR OUTPUT
SW 6 ON PUNCH
OFF
INCREASE T_e 3 %
SW 7 ON PUNCH
OFF
REDUCE T_e 200 TONS

Fig. 5. Flow chart, locomotive tonnage rating calculations program

Table II. Output Data from Locomotive Tonnage Rating Calculation

ROUTE: Akron TO: Hudson VIA:
Section: Entire Distance, Mile Post 15 to Mile Post 0
Locomotive Groups: 206–211 Tons Per Car: 40
Initial Speed: 35 mph At: Mile Post Location 14.85 Rails: Wet

Group	Tons	Cars	Distance	Mph	Time	Minimum Speed, Mph	At Distance	Identification
206	3,910	98	16.81	15.24	30.6	14.72	16.66	5
207	5,070	127	16.80	15.12	31.2	15.11	16.75	5
208	4,680	117	16.83	11.96	35.0	10.93	16.65	5
209	4,880	122	16.83	12.16	35.6	11.17	16.63	5
210	5,065	127	16.81	11.73	35.6	11.06	16.66	5
211	6,410	160	16.82	14.99	32.2	14.99	16.82	5

moving from one grade to another or may extend over more than one grade. In order to calculate acceleration, deceleration, and minimum speed accurately it is necessary to evaluate the average grade under the whole train at each instant.

The method of calculating tonnage ratings by use of the digital computer is as follows: Stored in tables on the memory drum of the computer are the locations of all grade lines, the grade or grade effect, and the maximum authorized speed ahead of each point of change in track data for a prescribed distance. This distance includes the principal grade under consideration and sufficient approach track to establish the normal approach speed. Also stored are the average gross weight per car and a table of tractive resistance versus speed.

Weight and drawbar pull data for each locomotive or group of locomotives are contained in a set of three cards to be read as each group is rated in turn, the least powerful being rated first. To start with, an estimated tonnage is assigned to the first locomotive. The computer, with this tonnage, calculates a run over the prescribed distance, noting the minimum speed of the train. If the minimum speed noted is too low, indicating trial tonnage too great, another run is made automatically with 200 tons less. If the minimum speed noted is not low enough, indicating insufficient tonnage, another run is made automatically with tonnage increased three per cent but not less than 25 tons. These calculations are repeated automatically until the minimum speed is sufficiently close to that desired, or the maximum permissible length of train has been exceeded.

Having thus calculated the tonnage rating for one group of locomotives, a card with this information is punched out and the data for the next group are read from the next three cards. As a trial tonnage for the second group, the computer will start with a tonnage proportioned to the rated tonnage of the preceding group in the ratio of the rated pulling powers of the two groups.

Fig. 5 shows the flow chart for the tonnage rating program. To obtain different amounts of output information, the calculations may be started at any one of three storage drum locations as follows:

LOCO (locomotive punch). An output card is punched only at the completion of the calculation of the tonnage rating for each locomotive.

RUN (run punch). An output card is punched at the end of each calculated run. The number of runs depends on how close the trial tonnage is to the rated tonnage.

TIME (time punch). An output card is punched at the end of each small time interval, Δt.

A portion of the tabulated output data from a tonnage rating calculation is shown in Table II. Items are as follows:

Column 1: By group is meant one or more classes of locomotive having sufficiently similar operating characteristics to be rated alike. There are seven classes in group 208.

Columns 2 and 3: Gross weight (tonnage) and number of cars.

Columns 4, 5, and 6: Distance, speed, and time at the point of data punch out. These items, when obtained at periodic short intervals of time by using the "time punch" routine of operating the computer, provide means of drawing a graph of the variables or of otherwise analyzing the train movement.

Columns 7 and 8: The minimum speed and the location at which it occurred.

Column 9: Punch identification, coded as follows:

0—Tonnage high
2—Tonnage low
5—Tonnage rating

Conclusions

The Pennsylvania Railroad Train Performance Calculator has proved to be a worthwhile and profitable investment. Because of the reduction in calculating time, many investigations and studies of ways of improving the train service, of improving the efficiency and economy of operation and some of the investigational work preparatory to the purchase of new locomotives have been accomplished which might not have been attempted without the calculator because of the number of man-hours required to make these calculations by hand.

It now appears that the use of the digital computer for this work makes possible a further and very much greater reduction in calculating time; train performance is calculated at rates varying between 700 and 2,000 miles of track per hour depending upon the number of station stops, the speed restrictions, and the grades.

Tonnage rating calculations on the digital computer require from one to five minutes per locomotive group depending on the length of track involved and the degree of accuracy desired. A tolerance of zero to minus 3 per cent of the exact rating has been found acceptable. To obtain the same results with comparable accuracy using the Train Performance Calculator and including the additional time required for manual preparation of data for each calculated run requires at least one man-day. To do the same work with slide rule and desk calculator, the time required would be prohibitive.

Because of the great reduction in time, the use of the digital computer has been found advantageous in calculating train performance and locomotive tonnage ratings. Being a general-purpose machine, it can be used also for a great variety of other data-processing work. It is expected that these newly developed uses may suggest other time-saving applications.

References

1. THE TRACTIVE RESISTANCE OF ELECTRIC LOCOMOTIVES AND CARS, W. J. Davis, Jr. *General Electric Review*, Schenectady, N. Y., Oct. 1926, pp. 18–24.

2. ELECTRICAL EQUIPMENT FOR RAILROAD DIESEL MOTIVE POWER (book), A. H. Candee. Westinghouse Electric Corporation, Pittsburgh, Pa., pp. 75–76.

3. A NEW TRAIN PERFORMANCE CALCULATOR, S. V. Smith. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 657–61.

Discussion

J. Stair, Jr. (Gibbs & Hill, Inc., New York, N. Y.): This subject is of vital interest to

engineers concerned with the economical operation of a railway.

Heretofore, such calculations had to be made by hand, involving much time, or the railway had to purchase an analog computer. Generally, therefore, the calculations were not made.

The great value of this paper lies in the fact that it presents a logical engineering analysis of the subject, and then shows how the calculations are made, more quickly than by any other method, with a digital computer—a machine that is available in the accounting department of most large railways.

J. E. Hogan: Since the paper was written, some additions have been made to the programs for the digital computer. In calculating train performance, means have been added for modifying the stored train data, i.e., changing the consist of a train (adding or cutting off cars or helper locomotives) at any intermediate points in a calculated run, with a minimum of additional computing time. By inserting two control cards and the cards for the modified train data at each desired location in the deck of track data cards, when the calculations reach each point of change of train consist, those changes which are indicated are made automatically in the stored train data.

Three additions have been made in the program for calculating tonnage ratings, as follows: 1. Tonnage ratings are automatically limited to keep the average acceleration during starting at values greater than 0.04 mile per hour per second; 2. means have been provided to calculate the minimum speeds obtained with specified tonnages, thus facilitating the checking of existing tonnage ratings; and 3. a procedure has been added for readily calculating tonnage ratings where helper locomotives are involved.

Overcurrent Protection in Large-Capacity Shipboard Electric Systems

J. R. COLE
ASSOCIATE MEMBER AIEE

IN all power systems the continuity of power service is dependent upon the characteristics of the devices that are employed in the distribution systems and the proper co-ordination of these devices. It is desirable to have these devices so arranged as to isolate any fault with a minimum disruption of power in the remainder of the power systems. The methods of accomplishing this on navy ships and the limitations of available protective devices were discussed in a previous paper.[1] Shipboard power plants have continued to become larger, thus resulting in higher available fault currents. This paper describes these larger systems and discusses the shipboard protective systems and devices employed.

Objective

It is the objective in navy shipboard distribution systems to have each unit of the equipment and all circuits protected from short-circuit currents and thermal overloads. The objective is to arrange and select the various protective devices to provide a complete co-ordinated protective system having the following characteristics:

1. High-speed clearing of all low-impedance faults.
2. Maximum continuity of service under fault conditions to be achieved by the selective operation of the various protective devices.
3. Maximum protection for electric apparatus and circuits under fault conditions by co-ordination of the thermal characteristics of the circuit or apparatus with the circuit-interrupting characteristics of the protective device.
4. Adequate interrupting capacity in all circuit-interrupting devices, or the provision of suitable backup protection, if the interrupting capacity of the device may be exceeded.
5. Adequate thermal rating in all of the various circuit protective and switching devices for operation under all service conditions.
6. Short-circuit current-carrying capacity of circuit breakers and bus-transfer equipment in excess of the maximum available short-circuit current within the maximum time limitations of circuit opening.

Selective Tripping

Selective tripping is accomplished by means of a time delay in the circuit breakers that is inversely proportional to current. This time delay is in the order of cycles. The breaker closest to the load may have an instantaneous trip. Each circuit breaker as it appears in the system proceeding from the load must then have a time-current band that will prevent it from initiating a trip until the circuit breaker in the system next closer to the load has sufficient time to trip and clear the fault. Fig. 1 is a schematic diagram of the distribution system on a large aircraft carrier, for which the time–current curves are given in Fig. 2.

Problems

In order to discuss some of the problems in this area, the following are defined:

1. *Maximum Available Average Rms Asymmetrical Short-Circuit Current.* This current which is the basis for interrupting rating of low-voltage (600-volt or less) air circuit breakers, is the average value of the maximum available rms short-circuit currents which occur in the three phases during the first half-cycle after inception of the short circuit. This current is based upon the maximum expected parallel generator capacity plus the contribution of the maximum induction and synchronous motor load connected to the system under this condition. In determining this current, all generators and significant circuit impedance up to the line terminals of the circuit breaker under study are considered.

2. *Maximum Available Maximum Rms Asymmetrical Short-Circuit Current.* This current is the basis for the interrupting rating of fuses, and the short-time current rating and mechanical strength of switches, cables, bus work, and circuit breakers. It is the rms current in the phase carrying the maximum asymmetrical current during the first half-cycle after inception of the fault. This current is based upon the maximum expected parallel generator capacity plus

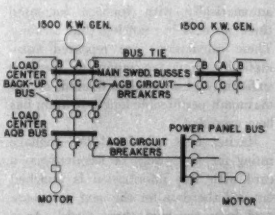


Fig. 1. Schematic diagram of distribution system on a large aircraft carrier

Paper 58-132, recommended by the AIEE Marine Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 16, 1957; made available for printing November 12, 1957.

J. R. Cole is with the U. S. Department of the Navy, Bureau of Ships, Washington, D. C.

Table I. Characteristics of Air Circuit Breaker, Navy Type ACB

Maximum Continuous Rating, Amperes	Navy Type ACB	Alternating Voltage, 60-Cycle	Maximum Interrupting Rating, Amperes	Short-Time Rating, Amperes	Size,* Inches (3-Pole) Height	Weight	Depth	Weight,* Pounds
250	250	500	20,000	15,000	17	12	9	100
640	640	500	45,000	25,000	24	13	13	217
1,600	1600	500	60,000	50,000	27	22	16	385
1,600	1600H	500	100,000	100,000	27	22	16	600
3,200	3200H	500	100,000	100,000	30½	30	25	700
4,000	4000H	500	100,000	100,000	30½	30	25	760

* Weight and dimensions are approximate.

the contribution of the maximum induction and synchronous motor load connected to the system under this condition. In determining this current all generator and significant circuit impedances up to the line terminals of the device under study are considered.

3. *Short-Time Current Rating of Circuit Breakers.* This is the maximum rms asymmetrical short-circuit current a circuit breaker will carry for the time duration that the breaker is intentionally delayed to produce the inverse time delay necessary for selective tripping.

4. *Interrupting Rating.* This is the maximum average rms asymmetrical short-circuit current that the circuit breaker will interrupt safely.

Inadequate Short-Time Current Rating

In order to have complete selectivity in a power system it is necessary that each protective device have adequate short-time current rating to withstand the fault current for the time delay required for selective operations. This means that the short-time current rating of the circuit breaker should be greater than the maximum available maximum rms asymmetrical short-circuit current at the point of application of the circuit breaker. If the protective device does not have this short-time current rating it is necessary to set the instantaneous trip at a maximum of 83% of the short-time current rating, see Table I. This is necessary so that the circuit breaker will trip instantly in case the fault current is above the short-time current rating and these prevent damage to the circuit breaker. The maximum setting of 83% of the rated short-time current is necessary since the calibration of the instantaneous trip ±20% accurate at this magnitude of current. Actually, it is normal practice to set this instantaneous trip at 80% of the short-time current rating of the breaker. Therefore, whenever the maximum available maximum rms asymmetrical short-circuit current at the point of application of a circuit breaker is greater than the circuit breaker short-time current rating it is necessary to set the circuit breaker instantaneous trip well below the fault current that may be expected, this limiting the selective tripping.

Inadequate Interrupting Capacity

If any circuit breaker in the system does not have sufficient interrupting capacity for faults that it may be called upon to interrupt, it must be provided with backup protection. This means that if the circuit breaker supplying a given load does not have an interrupting rating equal to or greater than the maximum available average rms asymmetrical short-circuit current, the breaker next in line proceeding from the load must have its instantaneous trip set at a maximum of 90% of the interrupting rating of the circuit breaker feeding the load. An example of this arrangement is the relationship between circuit breakers *D* and *F2* in breaker *F2* has a maximum interrupting capacity of 15,000 amperes. The maximum available average rms asymmetrical short-circuit current at the point of application of this circuit breaker is 56,000 amperes. Therefore, the circuit breaker *D* which has an interrupting capacity of 60,000 amperes must have its instantaneous trip set at 13,500 amperes to provide protection for the molded-case circuit breaker. In fact, in a great many instances large air circuit breakers have been provided on load centers only for the purpose of providing backup protection for a group of molded-case breakers. This instantaneous trip defeats selective tripping for all faults above 13,500 amperes.

Backup Protection

Where a molded-case circuit breaker is provided backup protection by the instantaneous trip of the circuit breaker feeding it, it is not expected that this backup protection will prevent damage to the molded-case circuit breaker in case of a fault. This backup protection will, however, prevent damage to any other circuit breaker in the same panel or switchgear unit. It may be necessary to replace the molded-case circuit breaker in the faulted circuit after the fault has been cleared by the backup circuit breaker.

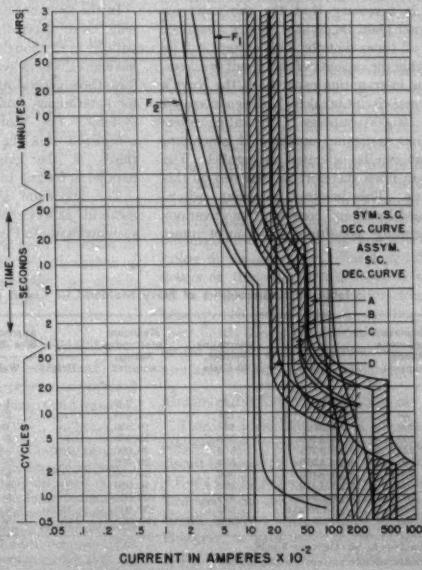


Fig. 2. Time-current characteristics of circuit breakers in Fig. 1

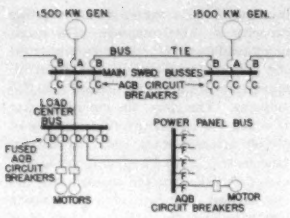


Fig. 3 (left). Schematic diagram of the distribution system on a large aircraft carrier provided with fused molded-case circuit breakers

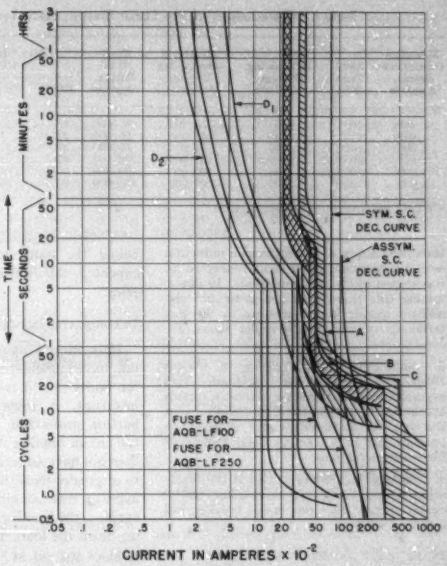


Fig. 4 (right). Time–current characteristics of circuit breakers in Fig. 3

Determining Magnitude of Short-Circuit Currents in the System

As shipboard power demands have increased, it has been necessary to install a greater number of larger generators aboard ship. In order to provide for continuity of service it has been considered necessary to provide bus ties in such a manner that it is physically possible to connect all ship's service generators on a ship in parallel.

It has been found that the most reliable as well as most flexible system can be attained by arranging the bus ties in such a manner that they form a closed loop if all bus tie breakers are closed. This arrangement, however, can result in potential fault currents well over the maximum interrupting capacity of the available circuit breakers if all generators are operated in parallel. Therefore it is necessary that the ship's service power plant be divided in such a manner at all times that the fault current will be within the capacity of the circuit breakers. The interrupting capacity of the Navy's largest circuit breakers has been increased to 100,000 amperes. This rating determines maximum numbers of generators that can be operated in parallel. The plant must therefore be designed on this basis. This involves balancing switchboard loads and consideration of various operating conditions. Personnel must also be indoctrinated to limit the number of generators that may be operated in parallel.

In the systems shown in Figs. 1 and 3 only three generators are operated in parallel even though each of these ships has eight 1,500-kw generator sets. Therefore, three generators in parallel are used for determining fault currents on these systems. The motor contribution is considered to be the contribution if two-thirds of the generator capacity in parallel is feeding induction motors.

The maximum available rms asymmetrical short-circuit currents of the systems shown in Figs. 1 and 2 are 73,000 average and 86,000 maximum at the generator distribution switchboard, 56,000 average and 62,000 maximum at a load center switchboard, and 55,000 average and 61,000 maximum at power panels.

Table II. Characteristics of Navy Molded-Case Circuit Breakers

Maximum Continuous Rating, Amperes	Navy Type	Alternating Voltage, 60-Cycle	Maximum Interrupting Rating, Amperes	Approximate Size, Inches			Weight Each, Pounds
				Height	Weight	Depth	
50	AQB-1*	125	5,000	3 1/2	1	2 1/2	1/2
50	AQB-5	125	5,000	4 1/4	3 1/4	3 1/2	2
200	AQB-10	125	10,000	5	4 1/4	4 1/2	4 1/2
100	AQB-A100	500	15,000	10	5	4 1/16	9 1/4
225	AQB-A225	500	20,000	15 1/2	9	4 13/16	26
250	AQB-A250	500	20,000	11	9	4 5/8	20 1/2
400	AQB-A400	500	30,000	15 1/2	9	4 5/8	34 1/2
600	AQB-A600	500	30,000	22	9	5 1/2	60
100	AQB-LF100†	500	100,000	10	5	4 1/16	12
225	AQB-LF225†	500	100,000	15 1/2	9	4 13/16	41
250	AQB-LF250†	500	100,000	12	9	4 5/8	34

* Single-pole, all others are 3-pole.
† Protected with current-limiting fuses.

Protective Devices

There are two basic types of circuit breakers employed on naval ships, the open-frame air circuit breaker, Navy type *ACB*; and molded-case circuit breakers.

Type ACB Circuit Breakers

Table I lists characteristics of the type-*ACB* circuit breaker. This type of breaker is provided with one or more of the following protective features integral with the breaker for protection of the distribution system.

1. Instantaneous overcurrent trip element.
2. Short-time-delay overcurrent trip element.
3. Long-time-delay overcurrent trip element.
4. Shunt trip.

In both the long-time-delay overcurrent trip and the short-time-delay overcurrent trip the tripping time is inversely proportional to the magnitude of the fault current. The curves in Figs. 2

and 4 show the time–current characteristics of type-*ACB* circuit breakers when applied to the systems shown in Figs. 1 and 3. Curves *A*, *B*, *C*, and *D* of Fig. 2, and curves *A*, *B*, and *C* of Fig. 4 represent type-*ACB* circuit breakers.

Molded-Case Circuit Breakers

Table II lists the characteristics of the molded-case circuit breakers employed by the Navy. Since this discussion is limited to the ship's 450-volt 60-cycle ship's service systems, only those with the type designations of *AQB-LF* will be discussed.

The type-*AQB* circuit breaker is provided with a long-time-delay thermal trip and an instantaneous trip. The long-time delay is set to trip in 25 seconds ±25% at a current of 600% of circuit breaker element rating, to trip in less than one hour at a current of 225% of element rating, and to trip in more than one hour at a current of 130% of element rating. There are two standard instantaneous trips for each element rating, one set to trip at 600–700% of the element rating and the other, in most cases, set to trip at 1,200–1,400% of element rating. Curve *F2* of Fig. 2 is a typical time–current characteristic curve for a type-*AQB* circuit breaker.

The type-*AQB-LF* circuit breaker is essentially a type-*AQB* circuit breaker equipped with current-limiting fuses. As has been previously pointed out, the low value of interrupting current of molded-case circuit breakers has in many cases required backup protection, thus negating selective tripping of the circuit breakers in the system. Fused molded-case circuit breakers eliminate this problem by providing adequate interrupting capacity. The fuse is mounted as an integral part of the circuit breaker and is arranged so that the blowing of a fuse trips all Table II assigns these breakers an interrupting rating of 100,000 amperes. Actually, when the average available rms asymmetrical short-circuit current at the point of application of these circuit breakers is any value between current-limiting fuses will limit the current to a value that will not damage the circuit breaker. The fuse should also have characteristics so that it will not burn out under fault currents that are within the interrupting capacity of the circuit breaker. Thus Navy specifications call for the time–current characteristics of the fuse in the fused circuit breaker to be such that the circuit breaker interrupts all overload and fault currents for all values of fault currents up to approximately 11,500 amperes for the Navy type-*AQB-LF 100* circuit breaker and approximately 16,000 amperes for the type-*AQB-LF 250* and *AQB-LF 225* circuit breakers without blowing, damaging, or causing any change in the time–current characteristics to occur in the fuse. This is being obtained on all fuse element sizes except the 15- and 25-ampere element. On these two sizes the crossover is at approximately 7,500 in lieu of 11,500. It has been necessary to use this lower crossover because the element of the circuit breaker is damaged when a fuse of higher let-through current is employed. Efforts are underway to increase this crossover current value to the same as is employed for the larger elements for the type-*AQB-LF 100* breaker.

Application of Protective Devices

The availability of circuit breakers with 100,000-ampere interrupting rating and a 100,000-ampere short-time current rating permits the use of larger ship's service generator capacity. The availability of these circuit breakers and the development of fused molded-case circuit breakers have resulted in a great degree of selective tripping. The large-capacity systems shown in Figs. 1 and 3 each consist of eight 1,500-kw (at 0.8 power factor) turbo-generators. Fault currents are based upon three generators in parallel. Figs. 2 and 4 are the time–current selective tripping curves for these systems showing the co-ordination of the circuit breakers. All circuit breakers in these systems are Navy type *ACB* or type *AQB*. Those designated as *A*, *B*, *C*, and *D* in Figs. 1 and 2, and *A*, *B*, and *C* in Figs. 3 and 4 are type *ACB*. Those designated as *D1* and *D2* in Figs. 3 and 4 are fused circuit breakers Navy type *AQB-LF* and those designated as *F1* and *F2* in Figs. 1 and 2 are the nonfused circuit breaker Navy type *AQB*. Thus the basic difference between these two systems is the use of the fused molded-case circuit breaker in Fig. 3. The following is a discussion of considerations in applying the circuit breakers in these systems.

Navy Type-ACB Circuit Breakers

The instantaneous overcurrent trip element is employed only when required for one of the following three reasons, otherwise it is eliminated.

1. Backup or cascade protection of a circuit breaker that has inadequate interrupting capacity as previously discussed herein. With development of the fused molded-case circuit breaker with an interrupting capacity of 100,000 amperes there is seldom need for an instantaneous trip for this purpose. The elimination of the instantaneous trip for this purpose improves the selective isolation of faults.

2. In the past instantaneous trip has been employed in certain applications because the short-time current rating of the circuit breaker was below the maximum available rms asymmetrical short-circuit current at the point of application of the circuit breaker. Therefore, in these instances it was necessary to provide the circuit breaker with an instantaneous trip to protect itself. With the development of the 100,000-ampere interrupting capacity circuit breaker with a short-time current rating of 100,000 this is no longer required.

3. The third case where instantaneous trip is employed is on generator circuit breakers where two or more generators may be operated in parallel. This instantaneous trip is to provide for isolation of faults on the generator side of the generator circuit breaker and is set so that the maximum available rms asymmetrical short-circuit current of one generator will not trip the instantaneous trip but so that faults of greater magnitudes trip the instantaneous element.

The short-time-delay overcurrent trip element of the type-*ACB* circuit breaker is the element that gives selective tripping Navy standards require four time–current bands for systems of the capacity under discussion. One band, which is the longest, is required for the generator circuit breaker, another for bus tie circuit breakers, another for feeder circuit breakers from the main generator and distribution switchboard, and a fourth for the feeder and backup circuit breakers on the load center switchboards. The time–current curves for these circuit breakers are represented in Figs. 2 and 4.

The function of the long-time-delay overcurrent trip is to provide a maximum degree of high-impedance fault protection in the feeder and bus tie circuit breakers. The long-time delay on the generator is to prevent excessive temperature rise of the generator. On the generator circuit breaker this device is set to pickup at a current value approximately equal to but not less than 150% of the full-load current rating of the generator. The long-time delays on the other circuit breakers are set in bands below the generator circuit breaker. However, there is overlap in these bands due to the limitations in the circuit breaker and the limited time–current area available for co-ordination.

Molded-Case Circuit Breakers

There is little likelihood of obtaining selectivity where two or more molded-case circuit breakers (type *AQB* or type *AQB-LF* or a combination thereof) are in series. Therefore, no more than two of these circuit breakers are permitted in series. However, to gain some degree of selectivity the circuit breaker feeding a group

of molded-case circuit breakers must always have a higher instantaneous trip setting than any circuit breaker which it feeds. Further, where two molded-case circuit breakers are in series, only the circuit breaker supplying power to the second circuit breaker is a fused circuit breaker since the one set of fuses will provide adcquate protection for both circuit breakers. Since the conventional rules are applied in matching molded-case element rating to loads, further selection of element sizes with respect to loads will not be discussed here. For complete selectivity it is desirable that the instantaneous trip on all molded-case circuit breakers be below the short-time pickup of the Navy type-*ACB* circuit breaker feeding it. However, where the fault current is greater than the instantaneous trip setting of the molded-case circuit breaker, and the instantaneous trip of the molded-case circuit breaker is greater than the pickup current setting of the type-*ACB* circuit breaker, selectivity exists between the two circuit breakers. Experience has shown that if the instantaneous trip of the molded-case circuit breaker is below the minimum fault current that may be expected under low-impedance faults, there is very little likelihood of the fault current being below the molded-case circuit-breaker instantaneous setting.

Contributions of Fused Molded-Case Circuit Breakers

In addition to the improved selectivity, the fused circuit breaker has contributed to a weight saving which is a very important item on shipboard. The fused circuit breaker has eliminated the need for type-*ACB* circuit breakers on the load center switchgear that were previously provided for backup protection for molded-case circuit breakers located on the same load center switchgear. With the development of the fused molded-case circuit breaker these molded-case breakers are connected directly to the load-center switchgear bus. It is estimated that 35 type-*ACB* circuit breakers have been eliminated on a large carrier with attendant savings in weight and cost.

Another contribution to weight and cost savings is in automatic bus-transfer equipment. Certain vital equipment and systems on navy ships must be provided with two or more power supplies with automatic bus transfer for transfer from one power supply to a second in case of power failure. The standard Navy contactor-type bus transfer has a short-time current rating of 20,000 amperes. Where the maximum available maximum rms asymmetrical fault current at the point of application of these bus transfers is above this 20,000 amperes, it has been necessary to provide bus transfers made up of electrically operated type-*ACB* circuit breakers of adequate short-time current rating. This has resulted in much larger, heavier, and costlier bus-transfer equipment than the contactor-type bus-transfer unit. Tests have shown that the fuses in the fused circuit breaker effectively limit the current times time (I^2T) fault currents that these contactor-type bus transfers may be called upon to carry to a value well below the thermal capacity of the contactors. Thus, where in the past it has been required to feed power panels via type-*ACB* circuit breaker bus-transfer equipment, the contactor-type bus transfer may be employed by using fused molded-case circuit breakers in the power panels.

It is estimated that 20,000 pounds are saved in switchgear and bus-transfer equipment on a large carrier by the use of the fused molded-case circuit breaker.

When fuses are burned out it is necessary to replace them. This is accomplished by removing a pull-out fuse assembly, replacing the fuses, and then replacing the assembly. The circuit breaker cannot be closed with a burned-out fuse, while the fuse assembly is removed, or while there is a fuse missing from one or more of the phases of the circuit breaker. The fuses are on the load side of the circuit breaker and cannot be removed without tripping the circuit breaker. There is therefore no hazard in removing the fuses and replacing them. Because of the high crossover between the overcurrent protection of the circuit breaker and the fuse, fuses are required to be replaced no more frequently than the molded-case circuit breakers where backup protection is necessary. The fused circuit breaker also eliminates any personnel hazard due to a molded-case circuit breaker trying to interrupt a fault current above its interrupting capacity.

Conclusions

The availability of the 100,000-ampere-interrupting-capacity open-frame air circuit breaker with the 100,000-ampere short-time current rating permits the use of larger ship's service generator capacity at 450 volts. The availability of these circuit breakers and the fused molded-case circuit breaker have resulted in a great degree of selective tripping in these power systems. The fused molded-case circuit breaker, in addition to providing greater selectivity on these large capacity systems, also contributes to a substantial savings in weight and cost.

References

1. Selective Tripping Arrangements for Protection of Ship's A-C Electric Distribution Systems. J. R. Cole. *AIEE Transactions*, vol. 67, pt. II, 1948, pp. 1256–62.

2. Co-ordination of Current-Limiting Fuses and Low-Voltage Circuit Breakers. M. S. Carlson, W. H. Edmunds. *Ibid.*, vol. 75, pt. III, Oct. 1956, pp. 1038–48.

Discussion

Richard R. McGee (General Electric Company, Plainville, Conn.): Mr. Cole's paper is a very clear and helpful presentation of the Bureau of Ships' objectives and philosophy of electric circuit protection. It is comforting to know that our armed services have made such a thorough analysis of the problem, are well abreast of developments in the electrical industry and even pushing some of them with decided vigor.

My first comment on the paper is that it is apparently Navy practice, and rightly so, to operate electric equipment to its full limit, with a narrower margin of safety than most nonmilitary applications. This is especially necessary on shipboard, I presume, where space and weight are at a premium.

My second comment is from the standpoint of the air-circuit-breaker application engineer. Most assuredly, the advent of the dependable high-interrupting-capacity, and at the same time current-limiting, fuse has opened up a wider field for the use of molded-case and switchboard-type air circuit breakers, to circuits of high available short-circuit current. Mr. Cole cites a specific saving of 10 tons in weight of one aircraft carrier by the use of this technique. Whether the combination breaker and specially built tripper-fuse, or the standard breaker and standard fuse, are used, a decided saving of size, weight, and cost can be realized, with no sacrifice of protection to circuit or personnel.

Where previously circuit breakers and fuses were considered irreconcilable competition and mortal enemies, now they are co-operating in a program in which each contributes vital elements and supplements the other's inherent limitations, for the common good.

The company with which I am associated has conducted an extensive development and testing program on current-limiting fuses, and their application with circuit breakers. The results of our tests and analysis of the fuse and molded-case breaker application for overload and short-circuit protection up to 100,000 amperes are summarized in a booklet, no. *GEA-6569A*,

which I will be glad to supply upon request.

Briefly, our tests have demonstrated that complete protection can be provided, single phasing avoided, and fuse replacements simplified by using standard breakers and standard current-limiting fuses, provided that properly co-ordinated ratings of breakers and fuses are used together. We have found that more than one rating of fuse must be used to secure proper co-ordination with a given frame size, eliminate nuisance breaker tripping, nuisance fuse blowing, or prevent damage to the calibration of the breaker thermal elements. The combinations of breaker and fuse ratings to achieve this result are listed in the booklet. The use of the right-rating current-limiting fuse in series with the molded-case breaker in the 50,000-ampere available circuit mentioned by Mr. Cole would prevent damage to the small circuit breaker and adjacent equipment.

With the availability of time-delay devices on magnetic trip elements of Navy type-*AQB* molded-case breakers, their application in selective systems becomes safe and feasible. Time delays of 3 and 5 cycles are available for the 400-ampere frame size, and 5 and 9 cycles will be available in the near future for ratings to 800 amperes in the *AQB-AT800*, based on the new General Electric type *TKM* breaker.

The new *AQB-A200* with ratings to 200 amperes, 500 volts, a-c, 10,000-ampere interrupting rating, will permit wider application and further weight savings. Its use with current-limiting fuses will handle 100,000-ampere available circuits.

Also available in molded-case breakers suitable for class *HI* applications per *MIL-C-17588* particularly for ships lighting and small power circuits are the single-pole type *ALB-1* and the 3-pole type *ALB-5* breakers, both in 10–50-ampere ratings, 125 volts, a-c/d-c with 5,000-ampere interrupting rating. The use of these with current-limiting fuses on ship circuits having higher available short-circuit currents might merit further study.

Our company pledges co-operation to the Navy Department in any measures to improve electrical protection and increase the efficiency of the fleet.

J. R. Cole: Even though it may be possible to obtain co-ordination between independent fuses and molded-case circuit breakers, to avoid single phasing, it is considered desirable to have them mounted in a single unit for shipboard applications. This arrangement allows for compactness, easy fuse replacement, and personnel safety when replacing fuses. It is anticipated that molded-case circuit breakers with cycle timers and the smaller size molded-case breakers will be employed aboard naval vessels when they become available.

Stability Study of a Third-Order Servomechanism With Multiplicative Feedback Control

Y. H. KU
FELLOW AIEE

C. F. CHEN
NONMEMBER AIEE

IN HIS PREVIOUS PAPERS[1-3] the first author has given a general analysis of nonlinear coupled circuits and applied the method to the analysis of servomechanisms with nonlinear feedback control. This paper reports the results of a stability study of a third-order servomechanism with multiplicative feedback control. The nonlinearities in the control system, and hence the differential equations representing the system, are introduced by the multiplication of certain quantities, or variables, of the system. For example, if the actuating signal e is multiplied by itself, one obtains e^2, which in turn introduces a nonlinear term in the final differential equation representing the feedback control system.

The choice of multiplicative feedback control is suggested by the favorable results obtained from a differential analyzer study of example I of reference 3. The third-order nonlinear differential equation representing the nonlinear control system, furthermore, is similar to van der Pol's equation. As the following study shows, a limit cycle in 3-dimensional phase space is obtained. This limit cycle in phase space serves as the separatrix between stable and unstable space trajectories.

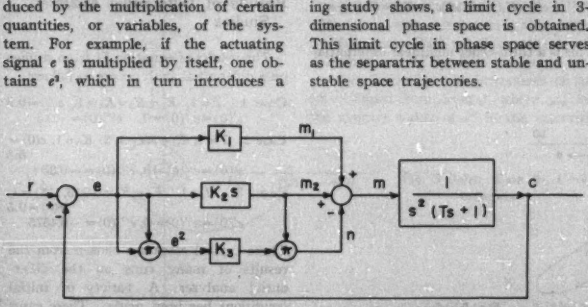


Fig. 1. Block diagram of third-order feedback control system

Nomenclature

$r(t)$ = reference input to system
$c(t)$ = output of system
$e(t) = r(t) - c(t)$ = actuating signal
$m(t)$ = total control effort in forward link
$m_1(t)$ = first component of control effort
$m_2(t)$ = second component of control effort
$n(t)$ = nonlinear component of control effort
s = complex variable in Laplace transforms
T = time constant
K_1 = gain constant
K_2 = gain constant
K_3 = gain constant

The sign π denotes multiplication, while the plus and minus signs denote addition and subtraction.

Differential Equations in Time Domain

In Fig. 1 the following differential equations may be written:

$$m(t) = Tc'''(t) + c''(t) \qquad (1)$$

Paper 57-1034, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted June 7, 1957; made available for printing December 23, 1957.

Y. H. Ku is with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., and C. F. Chen is with Christian Brothers College, Memphis, Tenn.

The first author wishes to acknowledge the encouragement and support of the National Science Foundation in the preparation of this paper. The second author wishes to thank the China Institute in America for awarding him a fellowship for the academic year 1956–1957 for undertaking this study as a partial fulfilment toward the degree of Master of Science in the Moore School of Electrical Engineering, University of Pennsylvania. To Drs. J. B. Brainerd and C. N. Weygandt and to M. E. Nelson, the authors wish to express their gratitude for help and encouragement. Credit is due to Morten Kenefsky for assistance in the differential analyzer study and to William Horrocks for supervising the making of the drawings used in this paper.

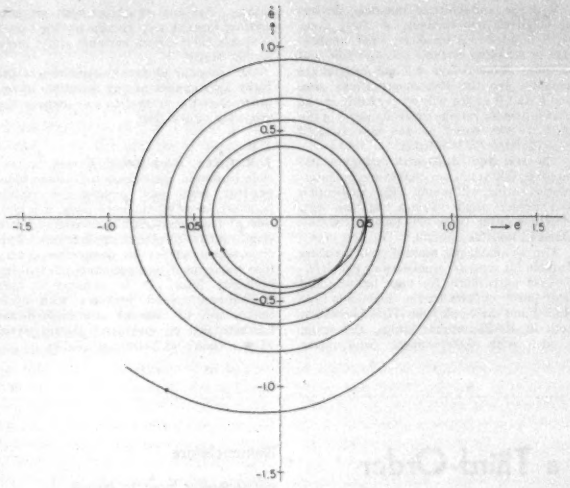


Fig. 2. Case 1: $e'-e$ trajectory

$$m_2(t) = K_2 e'(t) \tag{2}$$

where the prime denotes the first derivative with respect to time, and the double and triple primes denote the second and third derivatives respectively. The first component of the control effort is

$$m_1(t) = K_1 e(t) \tag{3}$$

The nonlinear component of the control effort is given by

$$n(t) = K_3 m_2(t) e^2(t) = K_2 K_3 e^2(t) e'(t) \tag{4}$$

The total control effort is

$$m(t) = m_1(t) + m_2(t) + n(t)$$
$$= K_1 e + K_2(1 - K_3 e^2) e' \tag{5}$$

By definition

$$r(t) - c(t) = e(t) \tag{6}$$

where $c(t)$ is equal to the feedback signal as shown in Fig. 1. It follows that

$$c(t) = r(t) - e(t) \tag{7}$$

$$c''(t) = r''(t) - e''(t) \tag{8}$$

$$c'''(t) = r'''(t) - e'''(t) \tag{9}$$

Substituting equations 8 and 9 in equation 1, equating the right-hand members of equations 1 and 5, and transposing terms gives

$$Te''' + e'' + K_2(1 - K_3 e^2) e' + K_1 e = Tr''' + r'' \tag{10}$$

This is a third-order nonlinear differential equation representing the control system with multiplicative nonlinear term $K_2(1-K_3e^2)e'$. For r''' and r'' equal to zero, equation 10 simplifies to

$$Te''' + e'' + K_2(1 - K_3 e^2) e' + K_1 e = 0 \tag{11}$$

Comparing equation 11 with van der Pol's equation, one notes that except for the third-order term Te''', the other terms are quite similar. However, in van der Pol's equation, the constant K_2 has a minus sign before it.

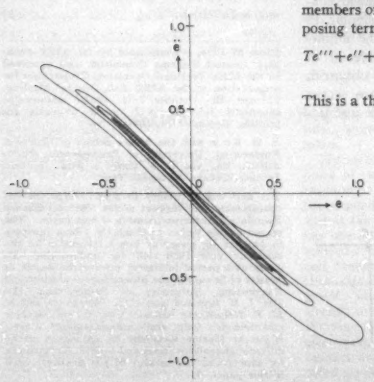


Fig. 3. Case 1: $e''-e$ trajectory

Phase-Space Equations

According to the method outlined in references 1, 3, and 4, the phase-space equations may be written as follows:

$$\frac{de''}{de'} = \frac{e'''}{e''} \tag{12}$$

$$\frac{de''}{de} = \frac{e'''}{e'} \tag{13}$$

where

$$e''' = (-1/T)[e'' + K_2(1 - K_3 e^2) e' + K_1 e] \tag{14}$$

In the present study, since a large number of values of K_1, K_2 and K_3 are to be inserted in the equation, it is thought number of values of K_1 and K_2 are to be inserted in the equation, it is thought advisable to use an analog computer to survey the effect of these values and then check the interesting sets by computation. In using an analog computer, moreover, one obtains a better insight into the stability of the system by observing and plotting the $e''-e'$, $e'-e$ or $e''-e$ trajectories of the system for different values of K_1, K_2 and K_3, as well as for different initial values of e and e'. Thus equations 11 and 12 may be obtained from analog computers in additions to plots of $e(t)$, $e'(t)$, $e''(t)$ and $e'''(t)$.

Results of Differential Analyzer Studies

The results of equation 11 from studies on a differential analyzer are as follows:

Case 1: $T=1$, $K_1=K_2=K_3=1$, $e(0)=0.5$
$e'(0)=e''(0)=0$, $e'''(0)=-0.5$

Case 2: $T=1$, $K_1=K_3=1/2$, $K_2=1$, $e(0)=0.5$
$e'(0)=e''(0)=0$, $e'''(0)=-0.25$

Case 3: $T=1$, $K_1=K_3=0.915$, $K_2=1$, $e(0)=0.5$
$e'(0)=e''(0)=0$, $e'''(0)=-0.4575$

These typical cases are chosen from the results of many runs on the differential analyzer. A variety of initial conditions has been made. These cases are reported here as they represent stable

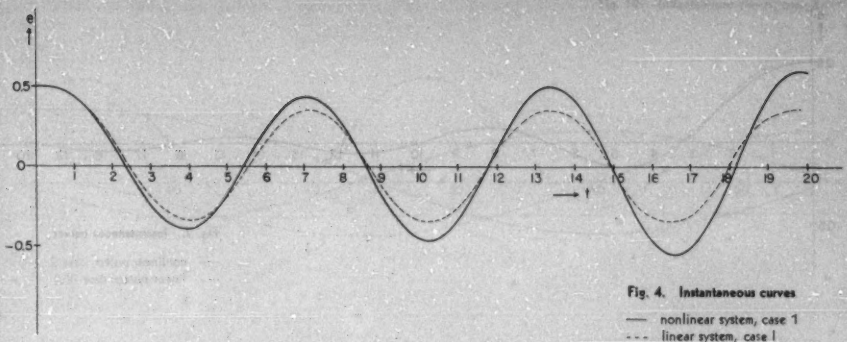


Fig. 4. Instantaneous curves

—— nonlinear system, case 1
- - - linear system, case 1

Table I. Calculation for Case 1

e	e'	e''	e'''	t
0.5000000	0	0	−0.5000000	0
0.4999944	−0.00025	−0.01125	−0.4885569	0.044
0.4999874	−0.00050	−0.01800	−0.4826124	0.061
0.4999688	−0.0010	−0.02685	−0.4723688	0.084
0.4999461	−0.0015	−0.03335	−0.4654711	0.101
0.4998886	−0.0025	−0.04335	−0.4546636	0.128
0.4997936	−0.0045	−0.059o	−0.4371186	0.168
0.4992286	−0.0095	−0.0843	−0.3927446	0.238
0.4985126	−0.0145	−0.1018	−0.3858376	0.292
0.4965326	−0.0245	−0.1298	−0.3483576	0.382
0.4942346	−0.0345	−0.1498	−0.3185576	0.454

Table II. Further Steps in Calculation

e	e'	e''	e'''	t
0.43	−0.18	−0.26	−0.023	1
0.12	−0.38	−0.08	0.334	2
−0.26	−0.30	0.27	0.276	3
−0.39	0.06	0.40	−0.061	4
−0.16	0.38	0.18	−0.390	5
0.23	0.37	−0.21	−0.376	6
0.43	0.03	−0.41	−0.044	7
0.24	−0.35	−0.25	0.339	8
−0.18	−0.41	0.17	0.406	9
−0.45	−0.09	0.44	0.081	10
−0.33	−0.33	0.35	−0.314	11
0.11	0.44	−0.08	−0.501	12
0.45	0.20	−0.44	−0.169	13
0.40	−0.27	−0.41	0.264	14
−0.03	−0.51	0	0.539	15
−0.46	−0.27	0.43	0.243	16
−0.50	0.21	0.49	−0.148	17
−0.08	0.55	0.10	−0.566	18
0.42	0.40	−0.40	−0.349	19
0.56	−0.10	−0.54	0.048	20

as well as unstable operations of the system under control.

Stable and Unstable Control

Case 1 represents an unstable control system. Fig. 2 shows the $e'-e$ trajectory, which starts at $e'=0$ and $e=0.5$ and spirals outward. Fig. 3 shows the $e''-e$ trajectory, which starts at $e''=0$ and $e=0.5$ and goes outward. The $e''-e'$ trajectory may be constructed by taking points from Figs. 2 and 3.

The instantaneous curves $e(t)$, $e'(t)$, and $e''(t)$ are obtained from the differential analyzer directly just as the aforementioned trajectories. It is believed that trajectories directly obtained from differential analyzer studies are here presented for the first time. Of course, these trajectories may be obtained either from actual experiments or from analog computers. Fig. 4 shows the instantaneous curve $e(t)$: thus for $t=20$ seconds the amplitude of oscillation is increased to such an extent that one may conclude that the system is unstable.

A step-by-step calculation is also made for this case. With small increments of Δv, where $v=e'$, the first ten steps are given in Table I. The increments of Δt are obtained from $\Delta e/e_{avg}'$, where e_{avg}' is the average value of e' in the interval corresponding to Δe. The corresponding values of e, e', e'' and e''' for further steps are given in Table II. From the data in Tables I and II, the slopes of the trajectories may be calculated and checked with those given in Figs. 2 and 3. In Fig. 4, the dotted-line curve shows $e(t)$ for a corresponding linear system with the nonlinear component of the control effort omitted.

While case 1 shows that the additional nonlinearity makes the system unstable, case 2 shows a very effective control system with the nonlinear control component present. The changes involved are the

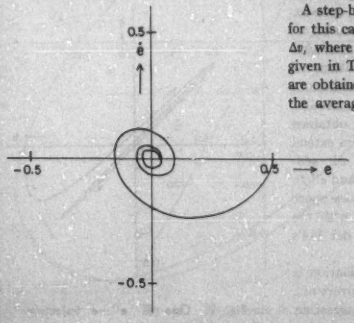


Fig. 5 (left). Case 2: $e'-e$ trajectory

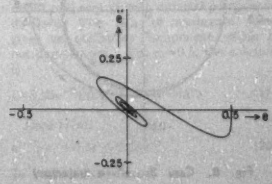


Fig. 6 (right). Case 2: $e''-e$ trajectory

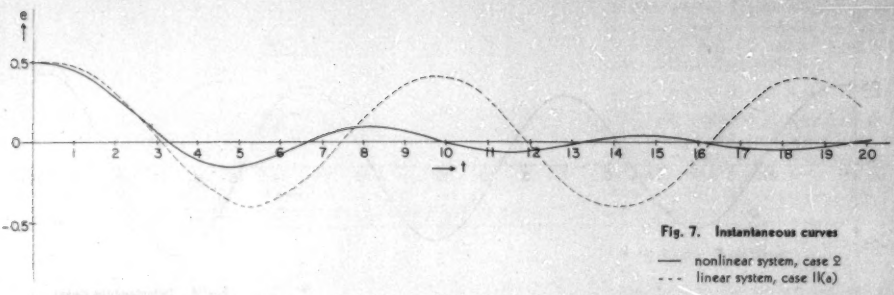


Fig. 7. Instantaneous curves

—— nonlinear system, case 2
- - - linear system, case II(a)

reduction of both K_1 and K_2 to 1/2. Fig. 5 shows the $e''-e$ trajectory which starts at $e'=0$ and $e=0.5$, and spirals to the origin of the plot. Fig. 6 shows the $e''-e$ trajectory which starts at $e''=0$ and $e=0.5$, crosses the horizontal axis, and spirals to the origin of the plot. Fig. 7 shows the instantaneous curve $e(t)$. It is seen that the actuating signal e, which is the difference between the reference input r and the output c, decreases rapidly to zero. This means that the output soon follows exactly the input without deviation. In Fig. 7, it is seen from the solid-line curve that the system is stable and the nonlinear damping very effective. The dashed-line curve represents $e(t)$ for case II(a) in the Appendix.

Preston has shown one optimum control system and one simplified control system.[5] In the simplified system a diode is used, and the linear component of the control effort is e'. In the stable control system here presented, the linear components of the control effort are $(e'+e/2)$, while the nonlinear component is $-e^2e'/2$.

Lewis reports a second-order nonlinear feedback control system with multiplicative term cc', where c is the output and c' its derivative.[6] The present third-order control system is hence a further extension of the second-order nonlinear feedback control system.

Self-Oscillations and Limit Cycle

Case 3 represents a peculiar system where self-oscillation occurs. Fig. 8 shows the $e'-e$ projection of a 3-dimensional space trajectory starting at $e=0.5$ and $e'=0$ and reaching a limit cycle in the form of a circle with $e_{max}'=\pm 0.35$ and $e_{max}=\pm 0.35$ approximately. Fig. 9 shows the $e''-e$ projection of the same space trajectory starting at $e=0.5$ and $e''=0$ and reaching a limit cycle with $e_{max}''=\pm 0.35$ and $e_{max}=\pm 0.35$ approximately. It is to be noted that the two projections reaching the form of limit cycles are really the projections of a 3-dimensional limit cycle in phase space, as are the projections of a general space trajectory discussed in reference 7. The authors believe that the extension of the limit cycle in a phase plane to the limit cycle in a phase space is a new contribution and hence of interest in the stability study of high-order control systems.

Fig. 10 shows the instantaneous curves of $e(t)$, $e'(t)$ and $e''(t)$. It is seen that after the half cycle of initial oscillation, all three curves show sustained oscillations and the period of a half-cycle is about 3 seconds. (The original curves obtained from the differential analyzer run extend to $t=20$.) The amplitudes of self-oscillation are 0.35 for $e(t)$, $e'(t)$ and $e''(t)$. It is of interest to compare the phase-space limit cycle, shown in Fig. 11, with the limit cycle obtained from van der Pol's equation.

In a control system, self-oscillation is undesirable. However, the discovery of a limit cycle in phase space indicates that the behavior of a high-order control system is predictable to a large extent, as the limit cycle forms a separatrix between the stable and unstable regions in phase-space.

In case 3, $K_1=K_2=0.915$ is determined from actual differential analyzer runs. Once this limit cycle is determined, it can be concluded that for K_1 and K_2 greater than 0.915, the system is unstable and that for K_1 and K_2 less than 0.915, the system is stable. Thus case 1 ($K_1=K_2=K_3=1$) falls within the unstable region, while case 2 ($K_1=K_2=0.5$, $K_3=1$) falls within the stable region.

Conclusions and Discussion

As a third-order transfer function (shown in Fig. 1) represents many practical systems under control, a nonlinear control system has been proposed and studied in this paper. The multiplicative feedback control represents in effect a computer which utilizes the actuating signal e to produce a nonlinear control effort m. It is seen that the total control effort has three components: two linear

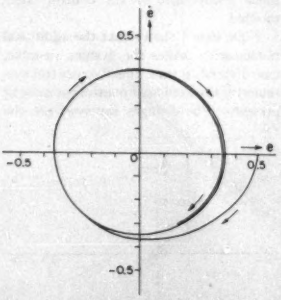


Fig. 8. Case 3: $e'-e$ trajectory

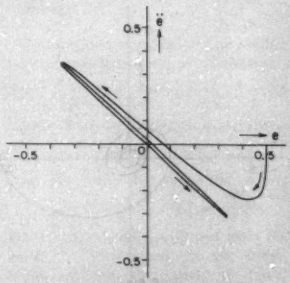


Fig. 9. Case 3: $e''-e$ trajectory

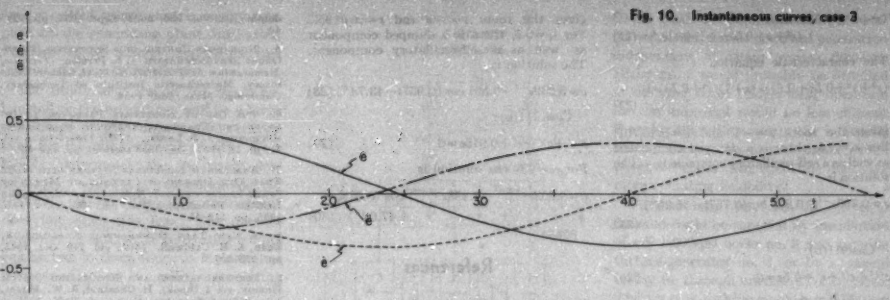


Fig. 10. Instantaneous curves, case 3

components and one nonlinear component. The first linear component is proportional to the actuating signal e. The second component is proportional to the first derivative of the actuating signal. The third and nonlinear component depends upon the product of e^2 and the second component. The third component in effect combines with the second component to form a nonlinear damping term in the final differential equation. Hence the whole control effort is a function of e, e', and the product of e^2 and e'.

In a linear system, the parameters are constants which determine the roots of the characteristic equation, and these roots in turn determine the stability of the system. In the Appendix, several linear systems are studied. It is seen that for cases I, II(a), and III(a), there are one real (negative) root and a pair of imaginary roots. Whether K_1 and K_2 are equal to 1.0, 0.5, or 0.915, the system has a self-oscillatory term when an initial value of the actuating signal is introduced.

Comparing the three cases in the nonlinear system with the corresponding cases in the linear system, it is seen that cases I, II(a), and III(a) in the linear system give rise to sustained self-oscillations, while only case 3 in the nonlinear system gives rise to sustained self-oscillations. Furthermore, case 2 in the nonlinear system gives a reasonably good control system which can be further improved. From the study of the limit cycle in phase space, it is to be noted that for K_1 and K_2 less than 0.915, there are many combinations of the constants that may make the control system effective and stable. Further study has been made on a Univac and the results are to be reported in a subsequent paper.

In conclusion, the authors wish to say that there are great possibilities in the nonlinear control of feedback systems. It is hoped that this report may stimulate further efforts in this field.

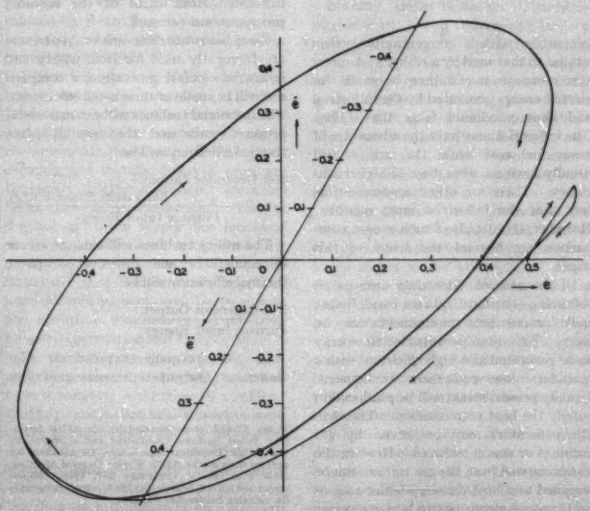


Fig. 11. Limit cycle in phase space, case 3

Appendix

If the nonlinearity is not present, equation 11 becomes

$$Te'''+e''+K_2e'+K_1e=0 \tag{15}$$

The foregoing three cases give the following linear differential equations.

Case I:

$$e'''+e''+e'+e=0 \tag{16}$$

Taking Laplace transforms gives

$$(s^3+s^2+s+1)E(s)=(s^2+s+1)e_0+(s+1)e_0' \tag{17}$$

where $E(s)$ is the Laplace transform of $e(t)$, e_0 and e_0' are the initial values of e and e' respectively. The characteristic equation

$$(s^3+s^2+s+1)=(s+1)(s^2+1)=0 \tag{18}$$

gives the roots $s=-1$ and $s=\pm j1$. The minus real root gives a damped component if $e_0 \neq 0$. The pair of conjugate roots gives an oscillatory component which is not damped out. For $e_0=0.5$, the solution is

$$e=0.25e^{-t}+0.3535\cos(t-45°) \tag{19}$$

Case II(a):

$$e'''+e''+0.5e'+0.5e=0 \tag{20}$$

In Laplace transforms,

$$(s^3+s^2+0.5s+0.5)E(s) = (s^2+s+0.5)e_0+(s+1)e_0' \quad (21)$$

The characteristic equation

$$(s^3+s^2+0.5s+0.5) = (s+1)(s^2+0.5) = 0 \quad (22)$$

gives the roots $s=-1$ and $s=\pm j0.707$. For $e_0=0.5$, there is a damped component as well as a self-oscillatory component. The solution is

$$e = 0.197e^{-t}+0.333\cos(0.707t-35.26°) \quad (23)$$

Case II(b):

$$e'''+e''+e'+0.5e=0 \quad (24)$$

For $e_0=0.5$, the solution is

$$e = 0.403e^{-0.682t}+0.326e^{-0.159t}\cos(0.861t-72.65°) \quad (25)$$

Case III(a):

$$e'''+e''+0.915e'+0.915e=0 \quad (26)$$

The characteristic equation

$$(s+1)(s^2+0.915) = 0 \quad (27)$$

gives the roots $s=-1$ and $s=\pm j0.957$. For $e_0=0.5$, there is a damped component as well as a self-oscillatory component. The solution is

$$e = 0.239e^{-t}+0.361\cos(0.957t-43.74°) \quad (28)$$

Case III(b):

$$e'''+e''+e'+0.915e=0 \quad (29)$$

For $e_0=0.5$, the solution is

$$e = 0.262e^{-0.904t}+0.353e^{-0.032t}\cos(0.978t-47.62°) \quad (30)$$

References

1. Analysis of Nonlinear Coupled Circuits, Y. H. Ku. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 626–31.

2. Analysis of Nonlinear Coupled Circuits—Part II, Y. H. Ku. *Ibid.*, vol. 74, pt. I, Sept. 1955, pp. 439–43.

3. Analysis of Servomechanisms with Nonlinear Feedback Control, Y. H. Ku. *Ibid.*, vol. 75, pt. II, 1956 (Jan. 1957 section), pp. 402–06.

4. A Method for Solving Third and Higher Order Nonlinear Differential Equations, Y. H. Ku. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 258, no. 3, Sept. 1953, pp. 229–44.

5. Nonlinear Control of a Saturating Third-Order Servomechanism, J. L. Preston. *Technical Memorandum 6897-TM-14*, Servomechanism Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., May 1954.

6. The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism, J. B. Lewis. *AIEE Transactions*, vol. 71, pt. II, 1952 (Jan. 1953 section), pp. 449–53.

7. Analysis of Nonlinear Systems with More Than One Degree of Freedom by Means of Space Trajectories, Y. H. Ku. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 259, no. 2, Feb. 1955, pp. 115–31.

8. A New Type Differential Analyzer, V. Bush, S. H. Caldwell. *Ibid.*, vol. 240, Oct. 1945, pp. 255–326.

9. Servomechanisms and Regulating System Design, vol. I (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc., New York, N. Y., 1955.

10. Control System Synthesis (book). J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

11. The Phase-Space Method for Analysis of Nonlinear Control Systems, Y. H. Ku. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 79, no. 8, 1957, pp. 1897–1903.

12. Analysis and Control of Nonlinear Systems (book), Y. H. Ku. The Ronald Press Company, New York, N. Y., 1958.

The Tie Between a Utility and an Industrial When the Industrial Has Generation

D. V. FAWCETT
ASSOCIATE MEMBER AIEE

LARGE industrial plants frequently generate all or a considerable amount of their electric energy requirements. The industrial may be remotely located so that a utility supply is not available. Alternatively, industrial generation may be more economical than power purchased from a utility. This can be the case if the industrial has water power or has low-cost by-product fuel available, or if its process heat requirements are large.

Today, in expanding existing plants and building new plants, it is often found desirable to have a utility tie even though the industrial operates its own generating station. This paper is intended to give a general outline of the considerations involved in operating industrial generation in parallel with a utility tie.

Generated Versus Purchased Power

An industrial with water power or low-cost by-product fuel available, such as steel-mill blast-furnace gas, may consider generation with a conventional system similar to that used by a utility. A direct cost comparison can then be made for electric energy generated by the industrial and that purchased from the utility. The industrial may have the advantage of lower fuel cost while the utility will usually have an advantage of higher efficiency. There are other economic factors that the industrial must consider. However, the details of such a cost comparison are beyond the scope of this paper.

In the case of industrials such as oil refineries, chemical plants, and paper mills, process heat requirements may be large. This often permits electric energy to be generated at a high efficiency with a relatively low equipment investment. Usually process steam will be produced to supply the heat requirements. The electric generators may be driven by gas turbines or steam turbines. Heat in the gas exhausted from the gas turbine can be supplied to a heat recovery boiler to generate process steam, or to a heat exchanger for heating the feedwater supply to a conventional fuel fired steam generator. Alternatively, the exhaust gas can be used as preheated combustion air for a steam generator since the oxygen content of the gas is almost as high as that in free air. If steam turbines are used, steam will be generated at a relatively high pressure, then supplied to the turbines and exhausted from them at the required process steam pressure.

Since steam-turbine-driven generators are frequently used for both utility and industrial electric generation, a comparison will be made of their usual efficiencies. The industrial turbines will exhaust useful process steam and the over-all plant thermal efficiency will be

$$\frac{\text{Net Generator Output} + \text{Exhausted Steam Energy}}{\text{Furnace Input Energy}}$$

The utility turbines will exhaust steam to condensers and the over-all plant thermal efficiency will be

$$\frac{\text{Net Generator Output}}{\text{Furnace Input Energy}}$$

In the foregoing expressions, Net Generator Output represents generator

Paper 57-642, recommended by the AIEE Industrial Power Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted March 25, 1957; made available for printing September 30, 1957.

D. V. Fawcett is with the Canadian Westinghouse Company, Ltd., Hamilton, Ont., Canada.

energy output minus electric energy supplied to the generating plant auxiliaries. Furnace Input Energy represents fuel burned in steam generation.

The utility heat loss in the steam exhausted to the condensers will be a large percentage of the furnace input energy even for the most efficient plants. This largely determines the relatively low efficiency which is characteristic of thermal plants. The industrial plant does not have this loss. The over-all efficiency of industrial generating plants can be two to three times as high as that corresponding to a utility generating plant.[1]

Also it should be realized that most of the equipment for the industrial plant, other than the turbine-generators and controls, is required anyway for the process steam generation. Of course the steam generation equipment must be designed for higher pressures. This does not result in too much additional investment. Similarly, the additional costs of manpower and other operating and maintenance expenses are relatively small.

The over-all industrial generating plant thermal efficiency using gas turbines can be almost as high as for the steam turbine application considered in the foregoing.

It should be noted that the foregoing expression for efficiency of the industrial plant applies on the following basis: When the generators produce the plant electrical kilowatt requirements, they will exhaust the steam flow (in say pounds per hour) required to supply the plant process steam requirements.

When the plant is built, the steam boilers and turbine generators can often be designed so that this requirement is met fairly closely for normal operation of the industrial processes. As the plant expands and installs different processes the electrical demand frequently increases faster than the steam demand. The industrial can often supply this increased electrical demand most economically by purchasing it from a utility. Even in the initial design of the plant, the most economical arrangement may be to supply the electrical requirements from both industrial generation and a utility tie.

On the other hand, some plants normally generate all their own power but have a stand-by connection to a utility.[2] Utility power is then available when some of the industrial generation is out of service. The cost to the industrial of a stand-by connection may be quite high. However, this arrangement may cost less than the installation of an equivalent amount of spare generating capacity.

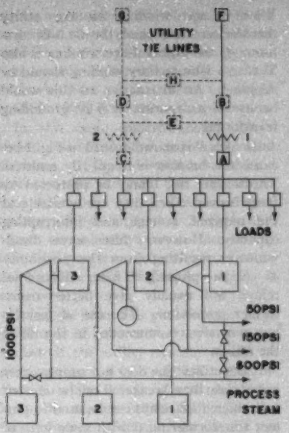


Fig. 1. Industrial plant generating system having tie lines to a utility

Many additional alternatives arise in obtaining the most economic arrangement of supplying the electrical and process steam requirements for various specific plants. Fig. 1 has been drawn to illustrate the way in which an industrial might be initially installed and later expanded to satisfy its varying process electrical and steam requirements.

Initially, steam boiler no. 1, turbine-generator no. 1, and perhaps one tie to a utility through circuit breaker *A*, would be installed. The boiler would supply say 600-psi (pounds per square inch) steam directly to a steam process header and also to turbine no. 1. Steam extraction from the turbine would supply, say, 150-psi process steam. Steam exhaust from the turbine would supply, say, 50-psi process steam.

If the electrical demand exceeds the steam demand, the additional electric energy could be generated by exhausting 50-psi steam to the atmosphere. However, power purchased from the utility is likely to cost less. If the 150-psi and 50-psi steam demand increases beyond that which can be supplied for the loading on generator no. 1, the additional steam can be supplied through the pressure-reducing valves shown connected between the steam process headers.

When a large increase in electrical demand results from a plant process expansion, a second turbine-generator no. 2 could be installed. This machine might have a condenser and also have steam extraction to either the 150-psi or 50-psi steam process headers. The condenser would permit electrical generation independent of steam process demand. Boiler no. 2 would probably be installed at the same time. Condensing operation for the industrial would be less efficient than for the utility unless the industrial has a source of low-cost by-product fuel. However, condensing operation does give more operating flexibility.

Further large increases in electrical demand could be supplied by installation of say 1,000-psi boiler no. 3 and topping turbine-generator no. 3, or by a second utility tie through breaker *C*, or by both. The topping turbine would use 1,000-psi steam and exhaust 600-psi steam. A pressure-reducing valve is shown connected in parallel with turbine no. 3. This permits use of steam from boiler no. 3 when turbine-generator no. 3 is out of service. Most efficient operation of the plant generating system will probably be obtained when generator no. 3 is operated at full load.

It should be pointed out that a tie between an industrial with generation and a utility can be a distinct advantage to the utility if it also purchases power on an interval demand basis. A small amount of power supplied from the industrial during a peak demand on the utility system may save the utility considerably on its purchased power cost. Also, the industrial may be able to provide stand-by capacity for local loads normally supplied by the utility. This requires that the industrial be able to adjust its load and generation at the same time the utility needs the power. Many industrial processes would not have this flexibility. Where such operation is possible, a mutually advantageous power contract can be negotiated.

There are a few cases in which a utility generating station and a large oil refinery have been built side by side and designed to operate as one over-all steam-electric generating system. In this arrangement, the utility operates the steam boilers and uses noncondensing turbine-generators exhausting steam to supply the refinery process steam requirements. The utility also supplies the refinery electric loads. The refinery returns condensate and supplies the fuel for steam generation. The over-all generating efficiency will be high. Since the refinery does not operate generators, this arrangement is outside the subject matter of this paper and accordingly will not be further considered.

The foregoing will illustrate the many arrangements possible for supplying the electrical and process steam requirements of an industrial.

Electric System Arrangement

We are all familiar with the many bussing and distribution circuit arrangements to be found in industrial electric systems. One of the major factors in determining bussing arrangements is the limitation of short-circuit momentary and interrupting duty requirements within the rating of the equipment being used. Other factors governing circuit arrangements are system flexibility, service continuity, voltage regulation, system efficiency, operation and maintenance costs, and initial investment. A paper by Lawrence and Thacker gives useful information on this subject.[3]

The most commonly used bussing arrangement is the single main bus. This has been shown in Fig. 1. Other arrangements could be used, as required, without changing the basic considerations of this paper.

Initially, perhaps one utility tie would be installed as indicated by breakers *A*, *B*, *F*. Transformer no. 1 would be required if the available utility supply voltage is above that for which the industrial generators could be designed. If the industrial generators could be designed to the utility supply voltage, then possible elimination of transformer no. 1 must be considered in the economic study which determines the industrial generating voltage.

The utility may require that the high-voltage winding of transformer no. 1 be Y-connected and grounded. This may permit cost savings in using graded transformer insulation. Also, if this winding is not grounded Y but is delta-connected, then a ground on the utility line, such as for a broken conductor, would trip breaker *F*. However there would be no fault current supplied from the industrial to trip breaker *A* and so the industrial generators would keep the utility line energized. This could be a safety hazard to the public. In addition, the possibility of arcing faults and high transient voltages may be undesirable. If this winding is Y-connected then the low-voltage winding should be delta-connected. This may be quite acceptable for the system in Fig. 1 if the industrial generators are grounded and at least one generator will be operating at all times.

However, some large industrials buying power at, say, 115 kv or higher may operate their generators at 13.8 kv and wish to use an intermediate voltage of, say, 34.5 kv to interconnect the generators and utility transformer and also to distribute to the various plant substations. In this case, the industrial may wish to use a Y-low-voltage winding on the utility transformer to ground the 34.5-kv system. If the high-voltage winding is also Y, then a delta tertiary winding should be supplied. An alternative to this would be use of a separate 34.5-kv grounding transformer.

Breaker *B* is shown dashed in Fig. 1 because this breaker is frequently omitted. Transformer no. 1 may be protected by fuses instead when they are available at the required voltage and interrupting capacity. However, fuses have disadvantages resulting from the possibility of single phasing.[4] Also, differential relays will usually give better transformer protection. The use of fuses is not generally recommended in the tie to the utility.

If the utility line does not supply other industrials, then breaker *B* can be omitted and differential relays can be used to protect transformer no. 1 with some form of remote tripping to assure that breaker *F* will trip. The differential relays may initiate remote tripping through pilot wire, power-line carrier, or one of the several available wireless communication systems, or a high-speed grounding switch may be closed on one of the phases of the utility line. This latter method should provide enough fault current to trip breaker *F* through its own protective relays.

If transformer no. 2 is installed and supplied from the same utility line, then high-voltage breakers are required for each transformer so that a fault in one will be cleared without tripping breaker *F*. This will permit the other transformer to continue to carry load.

If a second utility tie is installed to supply transformer no. 2 independently of transformer no. 1 as indicated by breakers *C*, *D*, *G*, then the same considerations will apply to breaker *D* as to breaker *B*.

Should the utility wish to join the two lines together, this connection can be made through circuit breaker *E* or breaker *H*. For an outage on one utility line, use of breaker *E* would permit both transformers to be fed from the remaining line. Use of breaker *H* permits the two lines to be tied through one breaker rather than three breakers. If breakers *B* and *D* are not installed, breaker *E* (same as breaker *H* in this case) would permit clearing a fault in one transformer with one of the utility lines being left in service.

Industrials operating generating systems without a utility tie may not have nor require lightning protection. This protection must be considered if a utility tie is installed since the utility system will be exposed to lightning. It is quite obvious that lightning protection will be required if the industrial is supplied directly from a utility overhead line. However, there are cases when protection will be required on the industrial main bus and distribution system even though the utility supply is through a transformer with lightning protection on its high-voltage terminals.

It should be emphasized at this point that the designers of industrial power systems should strive to their utmost for systems that are simple to operate and maintain. This will avoid unnecessary accidents and power interruptions.

System Stability

The electric system will likely have adequate stability if the industrial generates all its electric energy and is not tied with a utility system, or if the industrial purchases all its electric energy. When both generated and purchased energy is used, transient instability may occur for severe faults. The likelihood of instability is increased for the following conditions:

1. When the industrial and utility generation approach each other in magnitude.
2. When the utility tie has a high impedance construction and relatively low capacity rating.
3. When slow-speed relays and perhaps slow operating times are used.

It is recommended that large industrials investigate as to whether their systems will be subject to transient instability. This condition can usually be overcome by the use of high-speed relaying. Reference 5 describes the methods of making stability studies. Reference 6 refers to a study for an oil refinery.

Many large industrial plant loads may have an over-all power factor above 90%, without use of capacitors, if large synchronous motors are operated and most induction motors are loaded close to rating. This will apply to plants operating flow processes with a constant output. These industrials will often be capable of generating more reactive power than required for their plant load. If the utility can use this reactive, the industrial can supply it by increasing generator excitation, and at the same time continue to draw real power from the utility. This can result in a less costly power contact.

Another important result will be improved industrial voltage stability when the utility tie is lost. The loss of reactive demand by the tie will immediately aid in maintaining voltage in the industrial as

the increased kilowatt demand is thrown on the generators. This overcomes low voltage which would otherwise occur before the generator voltage regulators could take effect.

Load Shedding

Loss of the utility tie can result in a deficiency of industrial generating capacity. For this condition the industrial generators will slow down and not be able to maintain frequency. The frequency will decay at a varying rate as determined by the following:

1. Degree of overload as a percentage of generating capacity.
2. Inertia constants of the turbine-generators.
3. Reduction of steam-generating capacity because of reduced output from electric-motor-driven auxiliaries at their lower speed when operating on the decreased frequency.[7]
4. Reduction in degree of overload because of reduced output from process electric motor drives at the decreased frequency.

In item 1, the generating capacity is the total generation supplied by the generator turbines. Initially, it will be equal to the total generated power just before the utility tie is lost. After the turbine governors have responded, it will be the total power that can be generated within the limitations of turbine capability, governor settings, and the steam energy in the boilers.

In items 3 and 4, it should be noted that the horsepower requirements of various types of drives will change with the speed of the drive. Some drives may be regulated to draw constant horsepower. Most drives will not be regulated and as an approximation their horsepower requirements will vary directly as the frequency, the square of the frequency, or the cube of the frequency, depending on the type of drive.

Load must be dropped to restore frequency. If the industrial processes are not critically affected by low frequency and the decay is not too fast, the industrial station operators may be able to drop load and cope with the situation. For industrial loads which decrease appreciably with frequency, a small overload may largely be eliminated before the frequency drops too far. Also, a reduction in item 3 may not occur for industrials generating considerable process steam and also using steam-driven auxiliaries.

When the overload is a large percentage, published decay curves,[8] indicate that frequency may decrease quickly to relatively low values. Accordingly, a complete shutdown is likely to result. A typically severe overload would be for an industrial with generation capable of supplying only one-half its load. Loss of the utility tie would result in 100% overload on the generators.

In this case, underfrequency relays should be used for automatic load shedding. If the industrial has considerable nonessential load this can all be dropped at once at a predetermined frequency. However, it may be preferable to use several relays set at successively lower frequencies, each dropping one or more feeders until the frequency decay is arrested.

Even an industrial generating a large portion of its load can be severely overloaded, if the utility system should be split leaving utility load tied to the industrial. The industrial will require relaying to trip its utility tie breaker for this situation. Alternatively, if the industrial loses its generation it would not want this utility tie breaker tripped.

Several different relaying schemes have been used on utility systems for requirements such as these.[9] A reverse power relay may be used to trip breaker *A* for excess power flow to the utility. For industrials that supply power to the utility at times or for tie lines subject to power surges, a high relay setting may be required to avoid undesirable tripping. The power relay can be set to be more sensitive if it is used in conjunction with an underfrequency relay.

In many cases, the underfrequency relay only can be used with the power relay being omitted. The underfrequency relay should not trip during normal system frequency variations. Frequency is usually maintained within $\pm 1/2$ cps (cycles per second) on utility systems using steam turbine-generators. However, frequency may vary up to ± 2 cps on some systems using hydro-generators. If the frequency variation exceeds ± 1 cps, the power relay may be necessary. Also, the underfrequency relay should not trip in the event that the industrial loses generation. This will not be a problem if the industrial generation is small relative to the utility generation since in this case, the utility will be able to maintain near normal frequency.

The underfrequency relays used for load shedding should be set for a lower frequency or longer operating time so that the tie breaker can trip and drop the utility load before any industrial load shedding is done.

Loss of considerable industrial generation might result in a large overload on the utility tie. A power relay could be used to drop industrial load and sufficiently reduce the tie line loading to prevent tripping of the tie.

Some industrials purchasing power on an interval demand basis find it economically desirable to drop nonessential load instead of exceeding their normal peak on the utility tie. Power relays, or kilovolt-ampere relays when billing is on this basis, can be used either to trip or alarm. If the demand interval is long enough, an alarm may be sufficient to permit the station operator to drop the required load.

Utility Tie Line Reclosing

Utilities have found that continuity of service can be increased considerably by using fast reclosing on their overhead lines. In a typical reclosing arrangement, the breaker would open and reclose within 20 to 40 cycles after the fault occurred. Most line faults, except for broken conductors or insulators, are found to have cleared during the de-energized period involved in the foregoing breaker operation. If the fault fails to clear, the breaker will open a second time and reclose in possibly 20 to 30 seconds. After a third breaker operation similar to the second, the breaker will usually be locked open if the fault still remains.

Refer to Fig. 1 and assume that one tie is installed as indicated by breakers *A*, *B*, *F*. If breaker *F* is to be used for reclosing, then breaker *B* must be operated in conjunction with breaker *F*. Otherwise, the line would not be de-energized and the fault could not clear. It is not likely that the industrial would want its breaker *A* used for reclosing and so breaker *B* could not be omitted. Also, if breaker *B* was omitted, breaker *F* could not then easily distinguish between faults in the line and those in transformer no. 1 and reclosure on a fault in the transformer might occur.

Reclosing would benefit the industrial inasmuch as temporary tie line faults would not necessitate the industrial dropping load. However, reclosing between the utility and the industrial generators is not possible if the two systems swing out of phase appreciably during the reclosing interval. Second and third reclosures will usually not be possible.

Many of the factors previously considered in regard to transient stability and load shedding also are important in determining whether reclosing will be possible. During the interval that the tie is faulted and then de-energized, the industrial generators will be called upon to supply the power previously supplied by the tie. The generators will initially

slow down at a rate determined by the amount of additional power they must supply and their inertia constants. The reclosing interval is not long enough for changes in steam flow to occur. When the reclosing interval has elapsed the industrial voltage will then lag the utility voltage. If the angle of lag is less than the critical synchronizing angle of the tie line, the systems can be reclosed. The tie line would then transmit sufficient power to the industrial to pull the two systems back to the same frequency. If the tie line critical angle is exceeded, the systems would go out-of-step because the tie could not transmit sufficient synchronizing power to pull the systems together in frequency. The critical angle is determined by the relative values of tie line power flow just before the fault and the maximum power which can be transmitted over the tie.[10] This maximum power is a function of the tie line and transformer impedance and the voltage maintained at the ends of the tie line.

If single-pole reclosing is used, the tie would continue to supply power for one line-to-ground faults. The phase angle drift would usually be quite small for this condition.

Another consideration is the amount of inrush to the industrial generators and motors at the time of reclosing. This, of course, is a function of the angle of lag between the utility and industrial voltages. It is also governed by the generator and motor internal voltage and loading characteristics as well as the same factors which determine tie line maximum power. The inrush should not damage stator windings or result in excessive torques which will do mechanical damage to foundations, shafts, gears, or the various driving and driven equipment.

If fast reclosing is used (say 20 cycles) and the industrial generation is large relative to normal tie line power, the first reclosure may be possible. Large synchronous motors in the industrial plant will initially act as generators and help to prevent slow down of the system. On the other hand, if it is found that reclosing is not possible, then breaker B would not be equipped for it. Reclosing could still be used on breaker F to benefit other utility loads served from the same tie line. Breaker B would be tripped by suitable relaying and would lock open before breaker F recloses. Tripping of breaker B might then be used to initiate tripping of nonessential industrial feeders for load shedding purposes if a constant tie line loading is usually maintained.

Another alternative is available when reclosing on the industrial generators is not possible. The industrial main bus can be split with a tie breaker in between. Essential load and generators could be arranged on one bus section and the utility tie breaker and less essential load on the other. Tripping of breaker B would then be arranged to trip the bus-tie breaker. Breaker B would then reclose on load only. As indicated previously, fast reclosing can result in damaging inrush to large motors. Accordingly, these motors should be supplied from the generator bus section. Synchronizing facilities would be available across the bus tie breaker so that it could be manually closed if the utility reclosure was successful. Splitting the main bus in this manner requires that the normal load and available power on each bus should be balanced. The arrangement of loads and power sources on the bus as required in the foregoing and the use of a bus tie breaker may limit the flexibility of the electric system for future expansion.

If the second utility tie is added as indicated by breakers C and G in Fig. 1 and reclosing is to be used on both lines, then breakers B and D are required. Breakers D and G would reclose as a pair for faults on their line and B and F would reclose as a pair for faults on their line.

Reclosing may not be of much benefit for an industrial having its own generation and two utility tie lines because loss of one tie usually would not require load shedding. However, it may be important to the utility if the two lines are used as a major tie between two parts of their system.

The second line, with or without reclosing, gives much better service continuity. If one line faults and the other line is unaffected, then power flow to the industrial is only interrupted during the fault time. This may be 4 to 10 cycles if high-speed breakers and relaying are used. Reclosing of the faulted line is not likely to be necessary as far as the industrial is concerned. However, the faulted line probably could be reclosed without difficulty since the industrial and the other utility line will have stayed in step. When one line is out of service, the same considerations previously discussed for one tie line would apply regarding the possibility of satisfactory reclosing.

Voltage, Frequency, and Load Control

Load control is considered to involve the contol of flow of real and reactive power to the various loads. The net effect of an industrial having its own generation may be represented as a load supplied by the utility. Therefore load control can be considered with regard to actual plant loads within the industrial and also to the load transferred on the utility tie.

Since both voltage control and frequency control have a bearing on load control, these will be discussed first. The industrial will want to operate with the voltage essentially constant on its main bus. This voltage will be selected such that near rated voltage will be supplied to most of the plant motors for normal operation of the plant electric distribution system. The generators will usually have individual voltage regulators with reactive compensation. Bus voltage will then be regulated and the division of reactive power generation will be in accordance with the generator capabilities.

The utility tie transformer should be provided with a tap changer with an adequate range of voltage taps. The tap changer can be set so that the desired amount of reactive power is transferred over the tie line either to or from the industrial for the conditions of normal utility line voltage and normal industrial bus voltage. If both the reactive to be carried in the tie and the utility voltage will be fairly constant, then a no-load tap changer can be used. The industrial bus voltage can be varied somewhat by manual adjustment of the generator voltage regulators. This will maintain the required tie line reactive flow throughout slight variations of the utility voltage.

In the usual case of an important tie with the utility, an on-load tap changer will be required. The tap changer setting can then be changed under load to vary the tie line reactive flow as may be required and also to correct for utility voltage variation. Then the industrial bus can be maintained at proper voltage independently of tie line reactive flow.

An industrial having all its own generation and not tied with a utility system must control its frequency. Generators operating with speed control governors on their turbines will slow down somewhat as plant load kilowatt increases. This speed change will result in governor action to increase turbine steam flow and hence supply the increased load. The generator will now run slower and the frequency will be reduced correspondingly. The station operator will periodically check the frequency and if it has deviated much, will restore it to normal by changing the governor setting. The speed regulation from no-load to full-load will generally not exceed 4%. Some

plants will operate one turbine with an isochronous speed governor. This governor has essentially zero speed regulation and will take most of the kilowatt load swings itself while maintaining almost constant frequency.

Usually when industrial generation is operated in parallel with a utility, the latter will have considerably more generating capacity and will operate its system with relatively small frequency variations. Any tendency for the industrial generators to change speed would be immediately opposed by a flow of synchronizing power from the utility. Consequently all load kilowatt changes would be absorbed by the utility with little frequency change. Therefore the normal load regulating characteristic of the industrial turbine speed governor is not used.

If the utility system frequency should change, such as in a time of trouble, then industrial generators on speed control may change load considerably. A 1% drop in utility frequency may increase steam flow to the turbine so that the generator will pick up 25% kilowatt load or more. Isochronous speed governors cannot be used when tied in with a utility as extreme load changes would occur for slight frequency changes.

Many industrials with relatively constant load requirements will be able to maintain the required tie line kilowatt load by means of manual control of their turbine governors. Also they may manually control tie line reactive kilovolt-ampere load by generator voltage regulator adjustment or possibly on-load tap changer adjustment.

In other cases, the industrial operators may wish to have some combination of tie line kilowatt, reactive kilovolt-ampere, kilovolt-ampere, or power factor instruments which will alarm when prescribed operating limits are approached. This would be an aid for manual control.

Again, some industrials will require automatic tie line control. Devices similar to the foregoing can be used which will operate through some form of amplification to automatically adjust governors, voltage regulators, or tap changers as required. This is discussed in greater detail in other papers.[4,11]

When the industrial generators reach their upper load limits it may be desirable to have automatic tripping of some plant loads so that the prescribed tie line load limits are not exceeded.

The type and degree of load control requirements will vary from plant to plant depending on the type of processing and the terms of the power contract with the utility. There are many variations to be found in this regard and considerable attention should be devoted to predicting plant load requirements so that a power contract can be negotiated which will result in the minimum cost for purchased power.

Some industrials may be fortunate in having large motor loads whose outputs can be varied considerably without seriously affecting their processing. An example of this is to be found in pulp and paper mills. Often their grinders are driven by motors consuming a large amount of electric power. The grinder load can be varied manually or automatically in order to maintain constant loading of the utility tie.

General Relaying Considerations

In the past, many industrial power systems have been designed and operated with little or no consideration of protective relaying. In recent years, however, more attention has been given to industrial relaying requirements and the literature now contains several useful references.[1,12,13]

Induction overcurrent relays are generally used with industrial feeder breakers and these relays would also normally be applied to the utility tie breakers A and C of Fig. 1. Instantaneous trip attachments are applied within the limits of selective tripping requirements. Directional overcurrent relays are used when nondirectional relays would not operate selectively. Relay settings must be properly determined so that selective operation is obtained for all faults. For faults within the industrial system, the relays on breakers A and C will probably be required to co-ordinate with induction overcurrent relays applied as either primary or back-up protection on the breakers F and G (and breakers B, D, E, H if used).

Industrial generators and important transformers (this will include transformers no. 1 and no. 2 in Fig. 1) usually have differential relays for primary protection. Differential relays are frequently applied to protect the industrial main bus. Tie lines within the industrial distribution system often have pilot-wire relaying for high-speed protection.

The utility tie line may use distance, pilot-wire, or power-line carrier relaying for high-speed protection if the tie is considered to be an important circuit. High-speed relaying benefits the industrial because fault damage is reduced, the likelihood of transient instability is decreased, and shorter fault duration will result in fewer motors being tripped unnecessarily by their undervoltage protection.

Induction directional overcurrent relays with a low time setting will usually be applied to breakers A and C so that faults on the utility system are cleared from the industrial in a relatively short time. These relays must have time delay as required to co-ordinate with the first reclosing operation of utility tie line reclosing relays if they are used. For example, breaker A should not be tripped during the first reclosing operation of breaker B. However, if reclosing on the industrial is not possible, breaker B may not be installed. In this case breaker A should be tripped before the first reclosure of breaker F. The nondirectional relays used on breakers A and C as previously mentioned, should have sufficient time delay so that faults within the industrial system can be cleared without unnecessary tripping of either breaker.

If the utility tie is used primarily for stand-by purposes, there may be a danger after an outage on this line that the utility would energize it out of phase. For these circumstances, a sensitive reverse power relay can be used to trip breakers A and C on the exciting current transformers no. 1 and no. 2 would draw from the industrial upon loss of utility voltage.[9] Alternatively, relaying can be installed at the utility end of the line to prevent breaker F being closed if the line has voltage on it. This requires the utility to contact the industrial and arrange to have breaker A opened. Breaker F could then be closed and the two systems synchronized at breaker A.

As discussed previously, reverse power and underfrequency relays can be applied to trip breaker A if the utility system should be split leaving utility load tied to the industrial. Overload on the industrial system can then be eliminated by the use of underfrequency relays to drop load.

For loss of utility generation, the system frequency may drop considerably. The utility may then take power or it may continue to supply power to the industrial. If the frequency is likely to drop so low that the utility will take power or that the industrial plant equipment cannot operate satisfactorily, then an underfrequency relay can be used to trip breaker A. The use of reverse power and underfrequency relays to satisfy the requirements of a specific installation should be given particular attention.

In some cases the industrial may wish to use overfrequency relays to trip either the utility tie, or industrial loads such as large synchronous motors which may be adversely affected by high overfrequency.

High overfrequency would only occur on utility systems subject to losing a large block of their load at once. For this condition the industrial generators on speed control would be unloaded. Then if the utility tie is lost as a further result of the trouble on the utility system, the industrial generators would have all the load suddenly thrown on them. The industrial generating station may not be able to maintain operation for these violent changes required of the steam generating system.

If the utility tie is subject to single phasing because of blown fuses or broken conductors, the industrial generators may carry a dangerous amount of unbalanced loading. Phase balance relays are not likely to be sufficiently sensitive to protect the generators adequately. Sensitive negative sequence relays are available to give this protection.

Synchronizing facilities should be applied to all breakers, such as A and C, to be used for closing the two systems together.

Conclusions

The scope of this subject is so broad that all requirements which may vary from plant to plant cannot be discussed. Many general considerations have been outlined. The objective of this paper will be fulfilled if a further interest is stimulated in this subject. Each individual installation requires a detailed economic and engineering analysis to determine the most suitable arrangement of turbine-generators and utility ties and the various details of the over-all design.

References

1. Industrial Power Systems Handbook, Donald Beeman. McGraw-Hill Book Company, Inc., New York, N. Y., 1955, pp. 808–83.

2. Modern Electric Distribution System for an Oil Refinery, V. L. Nealy. *Electrical Engineering*, vol. 72, Nov. 1953, pp. 1008–11.

3. Electric Power for Refineries, R. F. Lawrence, H. B. Thacker. *Petroleum Processing*, New York, N. Y., vol. 8, Mar. 1953.

4. Problems Relating to Interconnections of Large Pulp and Paper Mills with Large Utility Power Systems, H. A. Rose, H.E. Springer. *AIEE Transactions*, vol. 73, pt. II Sept. 1954, pp. 205–16.

5. Electrical Transmission and Distribution Reference Book. Westinghouse Electric Corporation, Pittsburgh, Pa., 1950, pp. 433–95.

6. A Transient Stability and Voltage Study for a Modern Oil Refinery Distribution System, C. W. Boice, S. R. Durand, D. Dalasta. *AIEE Transactions*, vol. 75, pt. II, July 1957, pp. 159–64.

7. The Effect of Frequency Reduction on Plant Capacity and on System Operation, H. A. Bauman, G. R. Hahn, C. N. Metcalf. *Ibid.*, vol. 73, pt. III-B, 1954 (Feb. 1955 section), pp. 1632–37.

8. Application and Test of Frequency Relays for Load Shedding, L. L. Fountain, J. L. Blackburn. *Ibid.*, pp. 1660–64.

9. Automatic Load Shedding, AIEE Committee Report. *Ibid.*, vol. 74, pt. III, Dec. 1955, pp. 1143–46.

10. Protection for an Industrial Plant Connected to a Utility System with Reclosing Breakers, D. F. Langenwalter. *Distribution*, Schenectady, N. Y., vol. 15, no. 5, Nov. 1953, pp. 17–20.

11. Parallel Operation of Industrial Power Plants with Public Utilities, L. G. Levoy. *Ibid.*, pp. 10–16.

12. Protective Relaying for Industrial Electrical Systems, W. C. Woods. *Westinghouse Engineer*, Pittsburgh, Pa., vol. 16, no. 3, May 1956, pp. 92–96.

13. Electric Power Distribution for Industrial Plants. *AIEE Publication no. 952*, Oct. 1956, pp. 53–80.

Discussion

F. P. Brightman (General Electric Company, Schenectady, N. Y.): I agree with Mr. Fawcett's statement that it might be quite acceptable to ground the neutral of a system such as that shown in Fig. 1 by means of the generators only. In this case there are three generators and the general description of the system leads one to believe that it is reasonable to expect that at least one generator would always be in use, hence the system would always be grounded. However, since it is quite possible that the average industrial plant with local generation will not have a generator in use at all times, would it not be desirable to stress the advantages of having a grounded Y-connection on the industrial-plant side of the incoming transformers, to insure that the system would always be grounded.

The paper suggests that a reverse power relay be used to trip the incoming line breaker in the event that the utility system should be split apart leaving some portion of the power company's other-customer load connected to the industrial plant. My impression is that such a relay, when set slow enough to avoid false tripping on normal power swings, could not operate fast enough to save the industrial generating system from collapse, assuming the system is loaded to several times its normal capacity as is frequently the case. It would be helpful to have some specific information on relay settings and operating experience with such an installation.

The paper states that "underfrequency relays for load shedding should be set for a lower frequency or longer operating time so that the tie-breaker can trip before any load shedding is done." I interpret this to mean that one underfrequency relay is used for opening the tie-line circuit breaker as discussed in the preceding paragraph of the paper, and a second underfrequency relay is used for load shedding. Assuming that it is going to be necessary to shed load anyhow in order to prevent collapse of the industrial generating system, why not use one high-speed underfrequency relay to open the tie circuit breaker and shed load; i.e., do both jobs with one relay? Theoretically, this latter procedure may mean unnecessary shutdown of some of the feeders depending on total plant load at the moment. Practically, however, is the risk sufficient to warrant complicating the relaying system in an effort to minimize it?

As stated in the paper, two incoming-power lines operated in parallel will give the industrial plant much better service continuity than a single tie can. If automatic reclosing is provided on such lines, it is usually better to use slow reclosing, i.e., wait several seconds between successive attempts to re-establish the circuit that opened on a fault, than to use high-speed reclosing. Slow-speed reclosing gives the industrial plant's machines a chance to recover before being "hit" again with the low-voltage condition that results if the line is still faulted.

It is also very important to bear in mind when designing an industrial system, and particularly its utilization equipment, that the low voltage encountered during severe faults in either power system can cause unnecessary shutdown of motors unless suitable precautionary measures are taken. It may be permissible to let the motor starter contactor drop out in the case of most essential-service or critical-process induction motor drives, provided there is time-delay undervoltage protection to insure reclosing of the contactor when voltage is restored. However, there are many cases where it is desirable to keep the starter contactor closed during such disturbances, in order to minimize the "jolt" to the driven equipment when full voltage is restored. C. E. Quick, industrial power engineer of the Detroit Edison Company, reports that their investigations revealed that an outage of 30 cycles was permissible on low-voltage systems without spoiling tools or work in process, provided the motor line contactor was held in by maintaining control power by means of a voltage stabilizer.

It is also important to bear in mind that many synchronous motor drives can ride through 4 to 10 cycles of very low voltage without slipping a pole. This statement may seem a bit startling at first. The explanation lies in the fact that it is the angular displacement of the motor rotor and not the motor's speed at the end of the low-voltage period that is the determining factor. This angular displacement is the product of two factors, namely, the average motor speed in cycles per second for the time involved, and the time interval in seconds. To illustrate, the rotor of a motor which has slowed down to 58 cycles in 7 cycles will have been displaced only $(60-58)1/2(7/60)(360)=42$ degrees. This means that the starters for the synchronous motors on critical-process drives should be designed so that the motor is not disconnected from the line, or the field removed during such short-duration voltage disturbances by (1) dropping out of field contactors because of low voltage, or (2) incorrect operation of field removal relays.

I am assuming that the induction-type directional-overcurrent relays recommended for tripping the industrial end of the tie line in case of a fault in the utility system are the conventional inverse or very-inverse-time relays set to co-ordinate with the utilities' relays for faults in the utility system.

Sometimes, however, it is better to use a short-time relay, which will open the tie-line breaker at the industrial plant immediately without trying to co-ordinate with the utility system relays, and simultaneously disconnect nonessential load in the industrial plant. One small paper mill found it very desirable to do this because by so doing they could keep the paper machine running, whereas with conventional directional-overcurrent relaying, the mill was shut down at a cost of $700 or $800 in lost production every time a fault occurred in the utility system.

If an industrial plant has local generation, the importance of using high-speed underfrequency relays plus any other relays necessary to insure fast disconnection of the power company tie line and dropping of nonessential industrial plant load cannot be overestimated. It cost one large critical-process industry an estimated $250,000 to learn the value of such protection when there was a serious disturbance on the utility system.

As the author states, it is not general practice to use negative phase sequence or phase-balance relays for industrial-plant generator protection at the present time. It appears, however, that more consideration should be given to the use of such relays. It is rather doubtful that a phase-balance relay would be effective in preventing dangerous overheating of the generator rotor due to single-phasing of the tie between the industrial and the utility. Two cases are known in which small industrial-plant turbine generators were in trouble because of single-phasing of the tie line. In one case, a disconnecting switch in the tie-circuit was left open and the generator was severely damaged. In the other case, the line was for practical purposes single-phased by a high-resistance joint in a line connector, resulting in the generator's being considerably overheated but not damaged. The relay engineer of one large Eastern utility system recently stated that they were going to recommend the use of negative phase sequence relay protection for all industrial plant generators connected to their system.

When designing a protective relay system, it is important to keep in mind that while differential relays provide excellent first-line-of-defense protection for their particular zone, they will not operate for faults outside that zone. Therefore, since nothing is infallible, it is imperative that overcurrent relays be included for backup protection. Experience has conclusively demonstrated that the omission of the backup relays can be disastrous.

D. V. Fawcett: Mr. Brightman's discussion is a useful supplement to this paper. I agree with most of Mr. Brightman's comments and therefore will only refer to those requiring clarification.

Reference 9 of the paper refers to the use of the reverse power relay for load shedding and system separating. For satisfactory operation I believe this type of relay should be set relatively fast with its pickup high enough to prevent tripping on power swings. It then can be set fast enough to save the industrial generating system from collapse when the utility splits leaving utility loads on the industrial system. However, in the event that these utility loads could be below the pickup of the relay, and still sufficiently overload the industrial generation so that automatic tripping of the tie would be required, then of course this relay alone will not suffice. The paper discusses the use of an underfrequency relay either alone, or in conjunction with the power relay for these circumstances.

I do not feel there is any complication of relaying to use two underfrequency relays, each performing its one function, versus using one relay to perform both functions. In many industrial processes the improvement in continuity of service would economically justify the use of the two relays. In other cases only one relay would be chosen.

Reference 6 describes an oil refinery distribution system in which it was decided that the main contactors of 2,400-volt motor-starters should be held closed during system faults. This avoided the delay in restoring power to the motors that would otherwise have occurred during the closing time of the main contactors after voltage restoration to their closing coils. This reference reports that it was therefore possible to use normal torque motors whereas high-torque motors would have otherwise been required to reaccelerate the motor drives. Mr. Brightman has referred to a report by C. E. Quick. Reference 1 gives greater detail of investigations into the characteristics of various low-voltage industrial loads during voltage dips resulting from utility faults.

Mr. Brightman has pointed out that the ability of synchronous motors to ride through an interval of low voltage without pulling out of step is determined by the resulting angular displacement of the motor rotor. This angular displacement in electrical degrees can be calculated approximately as follows: Average motor slip (cps)×time interval (seconds)×360 electrical degrees per cycle.

It is hoped that this paper and the discussion will serve as a useful reference for the planning of future ties between industrial and utility generation.

Reference

1. Transients Trouble Suburban Industries, Charles E. Quick. *Electrical World*, New York, N. Y., Feb. 10, 1958, pp. 56–58.

Mobile Reflectoscopic Inspection of Railroad Car Axles on The Chesapeake and Ohio Railway

E. R. HAUER
NONMEMBER AIEE

C. M. ANGEL
NONMEMBER AIEE

IN 1952, The Chesapeake and Ohio Railway Company experienced a severe economic loss because of a freight train wreck. Investigation disclosed the derailment was caused by a broken axle, the failure starting from a crack in the journal. As a result of this accident, a study was begun by the mechanical department to determine a reliable and economical nondestructive test of railway car axles in service.

The Railway Company tests journal surfaces on all axles, of the plain bearing type in mounted wheels before they are placed in service and loose axles over their entire area before mounting, by the magnetic particle method and complies with all AAR (Association of American Railroads) requirements pertaining to axles. However, these methods will not detect internal defects in the axle, nor are they a practical method to detect flaws after the axles are placed in service. Inasmuch as the Railway Company has had considerable experience in the use of the Sperry Rail Service Company's reflectoscope (Figs. 1–3 briefly describe the principle of operation and show the cathode-ray oscilloscope indications of the reflectoscope) in the testing of steam locomotive axles and crank pins, a co-operative study was initiated with the Sperry Company to determine the feasibility of this device for testing car axles. A group of mounted axles with known flaws in the journals and wheel seats were selected for test. The journals were completely

Paper 58-238, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted June 5, 1958; made available for printing January 15, 1958.

E. R. Hauer and C. M. Angel are with The Chesapeake & Ohio Railway Company, the former in Richmond, Va., and the latter in Huntington, W. Va.

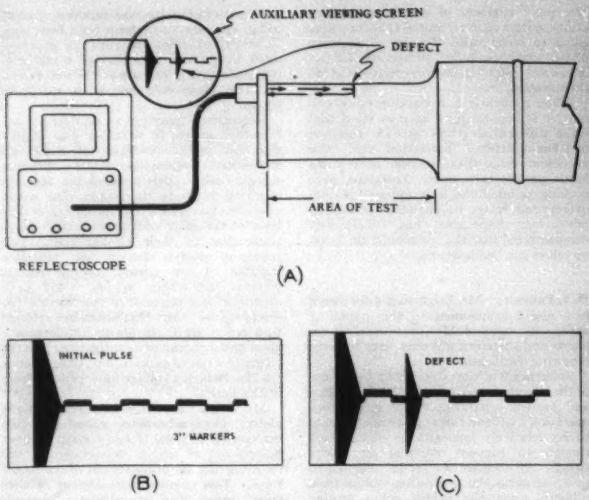


Fig. 1. Principle of operation

A—The Reflectoscope directs a pulse of sound energy into the axle, defects reflect a portion of this energy back to the search unit. The system of markers permits the operator to pinpoint the defect and its extent can be determined by moving the search unit over the surface of the end of the axle and watching the appearance of the indication on the viewing screen

B—A sound axle presents a clear base line. The simplicity (go–no go characteristic) of this inspection technique allows the use of a relatively unskilled mechanic as operator

C—A defect in the axle (surface or internal) will give the operator the above indication on his screen. Calibration of the instrument is rapid and simple and may be checked as frequently as desired on a sample bar of axle steel

Fig. 2. Initial pulse 3-inch markers on cathode-ray tube

Fig. 3. Pulse on cathode-ray tube locating failure in axle journal

covered with tape and an operator, unfamiliar with the condition of the axles, tested them with the reflectoscope; see Fig. 4. When defects were discovered, the operator was asked to outline the extent of the flaw with a pencil on the taped journal. It was found that all defects were detected and their locations were verified by magnetic particle testing. Further tests were conducted by installing axles, with known flaws, under empty and loaded cars and requiring the operators to locate the defective axles; see Figs. 5–6. In every case, the reflectoscope performed in a satisfactory manner; however, during this preliminary investigation, the reflectoscope was carried about in a wheelbarrow. This was not considered a practical service testing tool because of operator fatigue, low transport speed, and dependence on a long 110-volt a-c power cable.

To accomplish routine inspection of freight car axles in service, freight car repair yards were selected. In the yards there are personnel familiar with the AAR rules pertaining to axles and the necessary facilities to effect replacement should defective axles be discovered. In addition, cars find their way to these facilities frequently enough for this type of inspec-

Fig. 4 (left). Preliminary investigation. Locating cracks with reflectoscope in taped-axle journal

Fig. 5 (right). Preliminary investigation. Operator applying quartz crystal to end of axle with all parts in place

Fig. 6. Preliminary investigation. Quartz crystal placed on end of axle journal

Fig. 10. Midget motor automobile chassis being converted to a mobile reflectoscope carrier. Rear view

Fig. 13. Improved design of mobile reflectoscope carrier. Cathode-ray tube placed on right-hand fender giving operator better vision

Fig. 7. Wooden mock-up mobile carrier. Sliding seat side

Fig. 11. First design of mobile reflectoscope carrier. Instrument located on top of hood

Fig. 14. Front view of mobile reflectoscope carrier produced by Sperry Company

Fig. 8. Wooden mock-up mobile carrier. Drive engine side

Fig. 12. First design of mobile reflectoscope carrier. Operator placing quartz crystal on end of axle journal

Fig. 15. Side view of mobile reflectoscope carrier produced by Sperry Company

Fig. 9. Midget motor automobile chassis being converted to a mobile reflectoscope carrier. Side view

Fig. 16. Location of cathode-ray tube on mobile reflectoscope carrier produced by Sperry Company

tion and no delays would be caused to train and yard operation.

To be a workable tool, the reflectoscope, should be mounted on some type of mobile carrier, self-propelled, and equipped with an engine-driven generator to furnish electric current to operate the reflectoscope.

A survey of the problems involved to handle the reflectoscope in this manner was made. The mobile carrier would have to be close to the ground so the operator could reach the ends of the car journals with the reflectoscope crystal without leaving the seat of the mobile carrier. The vehicle would have to be easily maneuvered. The seat on which the operator sits would have to have a lateral sliding arrangement, so that when the journal ends were a considerable distance from the runway, the operator could run the machine with the seat pulled out.

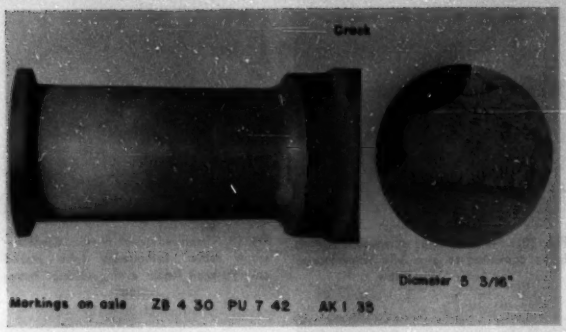


Fig. 17 (above). Defective axle

Mobile ultrasonic reflectoscope inspection
File: 700-2 (M-1-33)
Test Department: Metallurgical Laboratory, Huntington, W. Va.
Location: Newport News
Car No.: C & O 124774
Date Found: 2-11-55

Fig. 18 (left). Defective axle

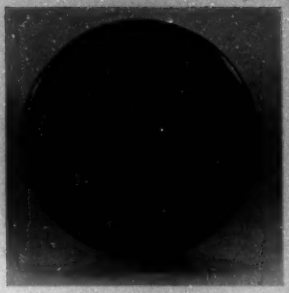


Fig. 20. Defective axle

keeping the mobile carrier on the runway and yet be able to place the reflectoscope crystal on the end of the car journal.

A powered mock-up of the proposed vehicle was constructed at the Huntington shops; see Figs. 7–8. Ten days after completion, the test department was evaluating its operational characteristics in the Huntington shop yard by checking miscellaneous cars. The reflectoscope showed indications of a cracked journal on

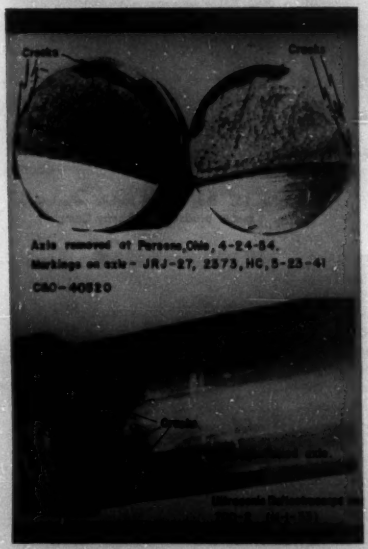


Fig. 19 (left). Defective axle

Fig. 21. Defective axle

Fig. 22. Defective axle

Table I. Chesapeake and Ohio Railway Company's 43 Months' Operation of Mobile Reflectoscope Units

Mobile Unit Location	No. in Service	Date Placed in Service	Total No. Cars Inspected	Cracked or Defective Journals Found			No. Cars Inspected Per Cracked or Defective Journal Found
				C&O	For-eign	Total	
Newport News, Va.	2	6/54; 12/54	33,357	38	30	68	490
Richmond, Va.	1	1/56	4,502	13	20	33	136
Clifton Forge, Va.	1	12/55	11,312	12	11	23	492
Handley, W. Va.	1	1/56	5,720	11	7	18	318
Peach Creek, W. Va.	1	1/56	5,854	4	3	7	836
Huntington, W. Va.	1	1/56	7,818	7	21	28	279
Russell, Ky.	2	1/56; 8/55	11,400	11	53	64	178
Ashland, Ky.	1	1/56	6,529	14	16	30	217
Stevens, Ky.	2	8/55; 1/56	11,866	7	11	18	659
Peru, Ind.	1	1/56	3,895	9	6	15	260
Columbus, Ohio	2	12/53; 12/54	46,006	70	28	98	469
Walbridge, Ohio	2	8/55; 8/55	23,988	10	19	29	827
Grand Rapids, Mich.	1	8/55	6,756	8	1	9	751
Flint, Mich.	1	12/55	4,758	7	12	19	250
Saginaw, Mich.	1	8/55	12,892	8	4	12	1074
Total or Average	20		196,653	229	242	471	418

Fig. 23. Defective axle

a revenue car. The wheels were removed from the car and the journal showed a long circumferential crack.

Further vehicle development centered around the use of a small automobile manufactured by the King Midget Automobile Company. Figs. 9–16 show the results of the continuing development carried out by the Railway Company, the automobile manufacturer, and the Sperry Rail Service Company.

In December 1953, the first of ten shop-built mobile units was placed in service at Columbus, Ohio. In November 1955, the Sperry Rail Service Company began offering complete mobile reflectoscope units. Since then, ten of these units have been added to the fleet.

Figs. 17–25 illustrate defective axles removed from revenue cars in service. In all cases, the defects were discovered by the mobile reflectoscope units. One axle was found with a crack that extended across approximately 50% of the cross-sectional area of the journal. The depth and location of these cracks was verified by breaking the journals in a one million pound testing machine.

Table I shows the results of 43 months of mobile reflectoscope unit operation. The 20 machines now in service have inspected a total of 196,653 freight cars, both Chesapeake and Ohio and foreign. Four hundred seventy-one cracked or defective axles have been found: 229 Chesapeake and Ohio and 242 foreign. This is an average of one defective axle for every 418 cars inspected. As will be noted in Table I, no definite trends have been established as to the frequency of defective axles at the various installations. Some locations have inspected a large number of cars, discovering only a few defects, while others have reflectoscoped a relatively few number of cars, but have found a large number of defective axles. This may be due in part to operator inexperience and to the fact that some locations are inspecting a high percentage of previously reflectoscoped axles. It is anticipated that as time goes by, operator proficiency will improve and, as a larger percentage of the total car freight fleet is inspected, more consistent results will be experienced.

Defective axles in service pose one of the most dangerous and pressing operating problems confronting railway management today. Ordinary inspection methods are ineffective in locating them and, invariably, failures occur in moving trains with disastrous results to rolling stock, roadbed, lading, and frequently to life. The mobile reflectoscope is a unique and effective answer to this aggravating problem. It provides a fast, inexpensive, and positive means of detecting incipient flaws in railway car axles before they can cause trouble, and it allows the mechanical forces to return cars to service with the assurance that the axles are safe. This device has proved itself in almost 4 years of service and has given The Chesapeake and Ohio Railway Company an outstanding operating advantage for today's high train speeds with the attendant severe loading of running gear components.

Fig. 24. Defective axle

Fig. 25. Defective axle

General Analysis and Stability Study of Finite Pulsed Feedback Systems

G. FARMANFARMA
STUDENT MEMBER AIEE

Synopsis: A finite pulsed system is defined as a system containing one or more samplers whose outputs are pulses of finite width and amplitude. In this report a method is presented for the exact analysis of finite pulsed feedback systems. The actual pulse width and the pulse shape are considered in its analytical treatment. The technique is based on the superposition principle, as well as some operational properties of the Laplace transform. It is applicable to a general linear system. The application of this method to systems with periodic samplers leads to a set of simultaneous first-order difference equations with constant coefficients. These equations can be solved for $C(z)$, the z-transform of the output at $t=nT$. The stability of the system can be determined from the roots of $C(z)$. It is found that the stability conditions are independent of the input but are functions of the pole-zero configurations of both the open- and closed-loop transfer functions, as well as the sampler characteristics. The technique is demonstrated by means of an example in the Appendix. Tables are also included to reduce the algebraic burden. These will be found useful in solution of second- and third-order systems. They can be extended whenever the need arises. The responses of an error pulsed second-order system to a step, ramp, and a sinusoidal input function are also plotted to show the effect of different pulse widths. These indicate that, in each case, as the pulse width h approaches the sampler period the output approaches the equivalent continuous system response. In a similar way it can be shown that as $h \to 0$ and the gain of the system is increased inversely as h, the output corresponds to the equivalent sampled-data system.

IN RECENT YEARS there has been a considerable amount of research in the field of sampled-data systems. Investigations by many authors has led to several techniques for the analysis and synthesis of such systems. The most important of these are the z-transform[1,2] and the modified z-transform.[3,4]

The term "sampled-data," as it is used in the literature today,[1,3,4] refers to a system with a sampling equipment which converts its input into a sequence of impulses. That is, it is assumed that the output of the sampler is a sequence of impulses having zero width, infinite height, and an area equal to the amplitude of the sampler input at the sampling instants. These impulses are then fed to the system either directly or through a hold circuit. With this assumption the z-transform and the modified z-transform may be used as mathematical tools to determine the system behavior. In practice this ideal condition cannot be realized and the z-transform approach becomes an approximate solution. However, it can be shown[5,6] that the approximation is good as long as the pulse width is small compared to the time constants of the systems.

The analytical treatment of the sampler as an impulse modulated device fails whenever the pulse width becomes appreciable. Hence a method must be developed which takes into account the actual pulse width and the pulse shape of the sampler output. Such a technique for the exact analysis of sampled systems with finite pulse width has been presented in earlier publications.[5-8] The mathematical details and their application to single sampler open-loop systems are discussed in references 5 and 6, while their extension to multiple sampler open-loop systems is carried out in references 7 and 8. It is the purpose of this report to supplement these and extend this technique to analysis of closed-loop systems. So far there has been no exact theoretical investigation of the finite pulsed closed-loop systems. Some experimental work has been carried out by Mr. Metzger[9] prior to this report. He used an analog computer to study the behavior of a second- and a third-order system with different pulse widths. These experimental results cannot be used to derive the stability criteria for the general system and cannot be correlated effectively in the absence of a theoretical analysis.

The study of finite pulsed systems is important in view of its applications. For example, in a telemetering system it is often necessary to use a single communication channel to transmit information about several independent variables. Under these conditions it becomes necessary to assign the communication link to each variable for a duration of time in rotation. This can be achieved, at the sending end, by a sampler with a pulse with h and period T. A similar synchronized sampler at the receiving end can then feed the received information to the correct data-reducing devices. Clearly such a system may be represented as a combination of several independent finite pulsed systems operating in parallel. Another application of the finite pulsed systems is in servomechanism, where the error is purposely pulsed and the pulse width is varied depending on the magnitude of the error. These systems are currently under experimental study. However, the principles outlined in this report are applicable to analysis of such servomechanisms.

In what follows, both the periodic and the nonperiodic cases are investigated. To avoid complications, the analysis is carried out for a system with unity feedback. But it will become evident that the same procedures can be used for the treatment of a general system with elements in the feedback path. An example is worked to show the effect of different pulse widths on the system response. It is shown that as h, the pulse width, approaches T, the sampler period, the system output approaches the continuous response. Also, as h approaches zero and the gain of the system is increased inversely as h, the system parameters approach those obtained through the use of the z-transform. The former limiting case is particularly interesting; it provides a totally different view of the continuous system. This could be advantageously employed in the synthesis of continuous servomechanisms.

The application of this method to a periodically pulsed system leads to a set of simultaneous first-order difference equations with constant coefficients. For an mth order system there are m such equations in m unknowns. These equations can be solved for $c(nT)$, the output at instants $t=nT$, by one of the many methods available. The z-transform method has been chosen in this report. This leads directly to an equation for $C(z)$, the z-transform of $c(nT)$. It is found that

---

Paper 58-240, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted September 30, 1957; made available for printing December 19, 1957.

G. FARMANFARMA is at the University of California, Berkeley, Calif.

The author wishes to express his sincere thanks and gratitude to Dr. E. I. Jury for his help and encouragement during the preparation of this report.

This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract no. *AF 18(600)-1521*. Reproduction in whole or in part is permitted for any purpose of the United States Government.

This material is part of the thesis to be submitted in partial fulfillment for the degree of Doctor of Philosophy at the University of California, Berkeley.

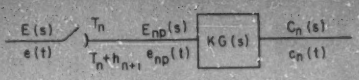
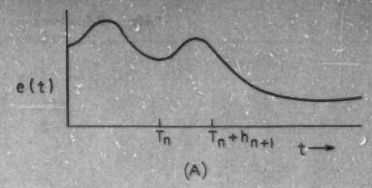
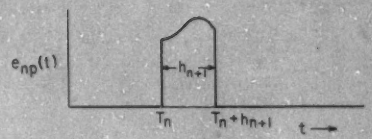


Fig. 1 (above, left). Finite pulsed system





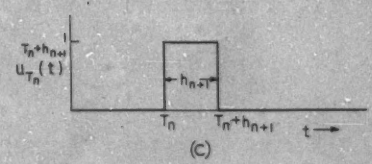


Fig. 2 (right). Input-output relation of a sampler

A—Sampler input
B—Sampler output
C—Pulse function

this method is more direct and results in equations whose forms are already familiar to the control engineer. The stability can be determined immediately from the poles of $C(s)$. For a stable system these must be within the unit circle. It can be shown that the stability of such systems is independent of the input and is only a function of the system parameters. The continuous response can also be obtained by further algebraic manipulations.

Mathematical Background

Consider the effect of a single pulse on the finite pulsed system shown in Fig. 1. Let $e(t)$, the sampler input, be the arbitrary time function shown in Fig. 2(A). Assume that the sampler is closed at the instant $t=T_n$, maintained there for a duration of h_{n+1} seconds, and then opened at $t=T_n+h_{n+1}$ seconds. If the sampler is ideal, the input to the system is the time function $e_{np}(t)$ shown in Fig. 2(B). Where

$$e_{np}(t)=e(t) \text{ for } T_n<t<T_n+h_{n+1} \quad (1)$$

$$e_{np}(t)=0 \text{ at all other times.}$$

Inspection shows that $e_{np}(t)$ may be written as the product of $e(t)$ and the function

$$u_{T_n}{}^{T_n+h_{n+1}}(t)$$

which, as shown in Fig. 2(C), is a pulse of unit amplitude stretching from the lower limit T_n to the higher limit T_n+h_{n+1} and is zero elsewhere. Therefore,

$$e_{np}(t)=e(t)\, u_{T_n}{}^{T_n+h_{n+1}}(t) \quad (2)$$

The Laplace transform of equation 2 may be evaluated by applying the complex convolution integral[3,6,10] and may be written as:

$$E_{np}(s)=\frac{1}{2\pi j}\int_\Gamma E(v)U_{T_n}{}^{T_n+h_{n+1}}(s-v)dv \quad (3)$$

where

$$E_{np}(s)=\mathcal{L}[e_{np}(t)] \quad (4)$$

$$E(v)=E(s)|_{s=v}=\mathcal{L}[e(t)]_{s=v} \quad (5)$$

$$U_{T_n}{}^{T_n+h_{n+1}}(s)=\mathcal{L}[u_{T_n}{}^{T_n+h_{n+1}}(t)] \quad (6)$$

and Γ is the path of integration enclosing all the poles of $E(s)$ to its left.[3,6,8] In general, $E(s)$ is a known function and $U_{T_n}{}^{T_n+h_{n+1}}(s)$ can be written as:

$$U_{T_n}{}^{T_n+h_{n+1}}(s)=\frac{e^{-T_n s}[1-e^{-h_{n+1}s}]}{s} \quad (7)$$

Thus $E_{np}(s)$, the Laplace transform of the sampler output, becomes:

$$E_{np}(s)=\frac{e^{-T_n s}}{2\pi j}\int_\Gamma E(v)e^{T_n v}\frac{1-e^{-h_{n+1}(s-v)}}{(s-v)}dv \quad (8)$$

Therefore, $E(s)$, the sampler input, is transformed according to equation 8 to give $E_{np}(s)$ the Laplace transform of its output. This will be referred to as p-transformation and denoted by the symbol P. This notation can now be used to write equation 8 in a simplified form as:

$$E_{np}(s)=P_{T_n}{}^{T_n+h_{n+1}}[E(s)]$$
$$=\mathcal{L}[e(t)u_{T_n}{}^{T_n+h_{n+1}}(t)] \quad (9)$$

where

T_n = instant at which sampler closes.
T_n+h_{n+1} = instant at which sampler opens.

This provides a convenient shorthand notation for expressing the relation between the input and output of the sampler. However, this notation is not employed for this purpose alone. In general, the symbol

$$P_x{}^y[F(s)]=\mathcal{L}[f(t)u_x{}^y(t)] \quad (10)$$

is used to denote the Laplace transform of the product of $f(t)$ and the pulse function $u_x{}^y(t)$ which is zero for $0<t<x$, has unit amplitude for the interval $x<t<y$ and is again zero for the duration $t>y$. In view of the above definition, it is easy to show that:

$$P_0{}^\infty[F(s)]=\mathcal{L}[f(t)\times 1]=F(s)$$
$$=\mathcal{L}[f(t)u_0{}^\infty(t)] \quad (11)$$

$$P_x{}^y[F(s)]=P_x{}^q[F(s)]+P_q{}^y[F(s)] \quad (12)$$
$$x<q<y$$

$$\mathcal{L}^{-1}\left\{P_x{}^y[F(s)]\right\}\begin{cases}=0 \text{ for } 0<t<x\\ =f(t) \text{ for } x<t<y\\ =0 \text{ for } t>y\end{cases} \quad (13)$$

and

$$P_x{}^y[e^{-Ts}F(s)]=e^{-Ts}P_{x-T}{}^{y}[F(s)] \quad (14)$$
$$T\leq x<y$$

In general, the lower limit x must be greater or equal to zero and the upper limit y can approach infinity. These properties of the symbol P are used throughout this report, and are an essential part of the analysis.

So far only a single pulse, i.e., a single closing and opening of the sampler, has been considered. To carry this basic step further, consider the effect of this pulse on the output of the system shown in Fig. 1. Let $C_n(s)$ denote the component of output due to this pulse. Assuming zero initial conditions $C_n(s)$ may be written, for all $t>0$, as:

$$C_n(s)=E_{np}(s)KG(s) \quad (15)$$

Some conclusions based on physical reasoning may be drawn regarding the behavior of $C_n(s)$ and hence its inverse $c_n(t)$. These can be proved mathematically and are as follows:

Table I. P-Transforms

No.	Function	Condition	$x=\infty$
1	$P_Z^{\infty}\left[\frac{a_0}{(s+\rho)}\right]$	$x\geq 0$	$e^{-\rho x}\left[\frac{a_0 e^{-\rho x}}{s+\rho}\right]$
2	$P_Z^{\infty}\left[\frac{a_0}{s^2}\right]$	$x\geq 0$	$e^{-\rho x}\left[\frac{a_0}{s^2}+\frac{a_0 x}{s}\right]$
3	$P_Z^{\infty}\left[\frac{a_1 s+a_0}{(s+\rho_1)(s+\rho_2)}\right]$	$x\geq 0$	$e^{-\rho_1 x}\left[\frac{a_0-a_1\rho_1}{\rho_2-\rho_1}\frac{e^{-\rho_1 x}}{s+\rho_1}+\frac{a_0-a_1\rho_2}{\rho_1-\rho_2}\frac{e^{-\rho_2 x}}{s+\rho_2}\right]$
4	$P_Z^{\infty}\left[\frac{a_1 s+a_0}{s^2+\omega^2}\right]$	$x\geq 0$	$e^{-\alpha x}\left[\frac{(a_1\cos\omega x+a_0/\omega\sin\omega x)s+(a_0\cos\omega x-a_1\omega\sin\omega x)}{(s^2+\omega^2)}\right]$
5	$P_Z^{\infty}\left[\frac{1-e^{-h(s+\alpha)}}{s+\alpha}\frac{a_0}{s+\rho}\right]$	$x\geq h$	$a_0 e^{-\rho x}\left[\frac{e^{-\rho x}}{\alpha-\rho}\frac{1-e^{-h(\alpha-\rho)}}{s+\rho}\right]$
6	$P_Z^{\infty}\left[\frac{1-e^{-h(s+\alpha)}}{s+\alpha}\frac{a_1 s+a_0}{(s+\rho_1)(s+\rho_2)}\right]$	$x\geq h$	$e^{-\rho_1 x}\left[\frac{a_0-a_1\rho_1}{(\rho_1-\rho_2)(\alpha-\rho_1)}\frac{e^{-\rho_1 x}(1-e^{-h(\alpha-\rho_1)})}{s+\rho_1}+\frac{a_0-a_1\rho_1}{(\rho_1-\rho_2)(\alpha-\rho_2)}\frac{e^{-\rho_2 x}(1-e^{-h(\alpha-\rho_2)})}{s+\rho_2}\right]$
7	$P_Z^{\infty}\left[\frac{1-e^{-hs}}{s}\frac{a_1 s+a_0}{s(s+\rho)}\right]$	$x\geq h$	$e^{-\rho x}\left[\frac{a_0 h}{\rho}\frac{1}{s}+\frac{a_0-a_1\rho}{\rho_2}\frac{e^{-\rho x}(1-e^{h\rho})}{s+\rho}\right]$
8	$P_Z^{\infty}\left[\left(\frac{1-e^{-hs}}{s^2}-\frac{he^{-hs}}{s}\right)\frac{a_1 s+a_0}{(s+\rho_1)(s+\rho_2)}\right]$	$x\geq h$	$e^{-\rho_1 x}\left\{\frac{a_0-a_1\rho_1}{\rho_1^2(\rho_1-\rho_2)}\frac{e^{-\rho_1 x}(1-e^{h\rho_1})+\rho_1 he^{-\rho_1(x-h)}}{s+\rho_1}+\frac{a_0-a_1\rho_2}{\rho_1^2(\rho_1-\rho_2)}\frac{e^{-\rho_2 x}(1-e^{h\rho_2})+\rho_2 he^{-\rho_2(x-h)}}{s+\rho_1}\right\}$
9	$P_Z^{\infty}\left[\frac{1-e^{-hs}}{s}\frac{a_1 s+a_0}{s(s+\rho_1)(s+\rho_2)}\right]$	$x\geq h$	$e^{-\rho_1 x}\left\{\frac{a_0 h}{\rho_1\rho_2}\frac{1}{s}+\frac{a_0-a_1\rho_1}{\rho_1^2(\rho_1-\rho_2)}\frac{e^{-\rho_1 x}(1-e^{h\rho_1})}{s+\rho_1}+\frac{a_0-a_1\rho_2}{\rho_1^2(\rho_1-\rho_2)}\frac{e^{-\rho_2 x}(1-e^{h\rho_2})}{s+\rho_2}\right\}$
10	$P_Z^{\infty}\left[\frac{1-e^{-h(s+\alpha)}}{s+\alpha}\frac{a_2 s^2+a_1 s+a_0}{(s+\rho_1)(s+\rho_2)\ s+\rho_3)}\right]$	$x\geq h$	$e^{-\rho_1 x}\left\{\frac{a_2\rho_1^2-a_1\rho_1+a_0}{(\rho_1-\rho_2)(\rho_1-\rho_3)(\alpha-\rho_1)}\times e^{-\rho_1 x}(1-e^{-h(\alpha-\rho_1)})+\frac{a_2\rho_2^2-a_1\rho_2+a_0}{(\rho_1-\rho_2)(\rho_2-\rho_3)(\alpha-\rho_2)}\times\frac{e^{-\rho_2 x}(1-e^{-h(\alpha-\rho_2)})}{s+\rho_2}+\frac{a_2\rho_3^2-a_1\rho_3+a_0}{(\rho_1-\rho_3)(\rho_2-\rho_3)(\alpha-\rho_3)}\times\frac{e^{-\rho_3 x}(1-e^{-h(\alpha-\rho_3)})}{s+\rho_3}\right\}$

1. $c_n(t)$ is zero for the interval $0<t<T_n$, because the input $e_{np}(t)$ is zero during this interval. (16A)

2. For the duration of h_{n+1}, i.e., for $T_n<t<T_n+h_{n+1}$, the poles of $C_n(s)$ are the same as the poles of $E_{np}(s)\,KG(s)$. This means $c_n(t)$ will rise or decay with time constants determined by the poles of $E_{np}(s)$ as well as those of $KG(s)$. The reason for this is the fact that during this interval $e_{np}(t)$ is acting as the forcing function. (16B)

3. For $t>T_n+h_{n+1}$ the poles of $C_n(s)$ are the same as the poles of $KG(s)$ alone. This is because the opening of the sampler results in zero input. The response is then the same as when the system is given some initial conditions and allowed to decay. Hence, during this interval the rise or decay of $c_n(t)$ is governed entirely by the poles of $KG(s)$. (16C)

A simple example will illustrate the above-mentioned characteristics. Referring to Fig. 1, let

$$E(s)=\frac{1}{s+\alpha}$$

and

$$KG(s)=\frac{1}{s+\rho}\ (\text{where } \alpha\neq\rho) \tag{17}$$

Assume that the sampler is closed at a time $t=T_0$, maintained there for the duration of h_1 seconds, and then opened at time $t=T_0+h_1$ seconds. $E_{0p}(s)$, the Laplace transform of input pulse, may be written as:

$$E_{0p}(s)=P_{T_0}^{T_0+h_1}\left[\frac{1}{s+\alpha}\right]=\frac{e^{-T_0(s+\alpha)}[1-e^{-h_1(s+\alpha)}]}{(s+\alpha)} \tag{18}$$

Therefore, $C_0(s)$, the component of output due to this pulse, is:

$$C_0(s)=e^{-T_0(s+\alpha)}\frac{1-e^{-h_1(s+\alpha)}}{(s+\alpha)(s+\rho)} \tag{19}$$

The inverse of equation 19 can be written as:

$$c_0(t)=\left\{d(-T_0)\left[\frac{e^{-T_0\alpha}}{\alpha-\rho}(e^{-\rho t}-e^{-\alpha t})\right]-d[-(T_0+h_1)]\left[\frac{e^{-(T_0+h_1)\alpha}}{\alpha-\rho}(e^{-\rho t}-e^{-\alpha t})\right]\right\} \tag{20}$$

where $d(-T_0)$ is the notation used to represent a delay of T_0 seconds along the time axis due to the $e^{-T_0 s}$ multiplier. Equation 20 may be simplified and expressed in terms of three separate components. These are:

$$c_{00}(t)=0 \text{ valid for } 0<t<T_0 \tag{21}$$

$$c_{01}(t)=\frac{e^{-T_0\alpha}}{\alpha-\rho}[e^{-\rho(t-T_0)}-e^{-\alpha(t-T_0)}] \text{ for } T_0<t<T_0+h_1 \tag{22}$$

and

$$c_{02}(t)=\left[\frac{e^{-T_0\alpha}(e^{-\rho h_1}-e^{-\alpha h_1})}{(\alpha-\rho)}e^{-\rho[t-(T_0+h_1)]}\right] \text{ for all } t>T_0+h_1 \tag{23}$$

Equations 21, 22, and 23 verify the results already mentioned. In particular, equation 23 shows that for all $t>T_0+h_1$, i.e., when the sampler is open, the rise or decay of the output is determined only by the value of ρ, the pole of $KG(s)$, and is independent of α, the pole of $E_{np}(s)$.

Equation 19 expresses $C_0(s)$, the Laplace transform of the output for all times $t>0$. However, it is seen from equations 22 and 23 that the transform of the output simplifies greatly over specified regions of time. For example, the Laplace transform of equation 23 is:

$$C_{02}(s)=e^{-(T_0+h_1)s}\left[\frac{e^{-T_0\alpha}(e^{-\rho h_1}-e^{-\alpha h_1})}{(\alpha-\rho)}\right]\times\frac{1}{s+\rho}=\mathcal{L}\,[c_{02}(t)] \tag{24}$$

which is much simpler to handle than equation 19 and represents the correct output for all $t>T_0+h_1$. For this reason, it is often advantageous to obtain and utilize these piecewise transforms. However, this must be done directly in s-plane if it is to be of much practical value. The P-notation and some of its properties outlined previously provide a convenient method through which this can be achieved. To show this, the Laplace transform of equation 23 will be derived directly from equation 19. Using relation 11, $C_0(s)$ may be written as:

$$C_0(s)=P_0^{\infty}[C_0(s)]=P_0^{\infty}\times\left[\frac{e^{-T_0(s+\alpha)}}{(s+\alpha)(s+\rho)}-\frac{e^{-(T_0+h_1)(s+\alpha)}}{(s+\alpha)(s+\rho)}\right] \tag{25}$$

Equation 12 may be utilized to write:

$$C_0(s)=P_0^{T_0}[C_0(s)]+P_{T_0}^{T_0+h_1}[C_0(s)]+P_{T_0+h_1}^{\infty}[C_0(s)] \tag{26}$$

The three components of equation 26 represent the transforms of equations 21, 22, and 23, respectively. Consider the third component. Using relations 14 and 25, this can be written as:[7,8]

$$P_{T_0+h_1}^{\infty}[C_0(s)]=\left\{\frac{e^{-T_0(s+\alpha)}}{2\pi j}\times\int_\Gamma\frac{1}{(v+\alpha)(v+\rho)}\frac{e^{-h_1(s-v)}}{s-v}\,dv-\frac{e^{-(T_0+h_1)(s+\alpha)}}{2\pi j}\int_\Gamma\frac{1}{(v+\alpha)(v+\rho)}\times\frac{1}{s-v}\,dv\right. \tag{27}$$

which reduces to:

$$P_{T_0+h_1}{}^{\infty}[C_0(s)] = e^{-(T_0+h_1)s} \times \left\{ e^{-T_0\alpha}\left[\frac{e^{-h_1\alpha}}{\rho-\alpha}\frac{1}{s+\alpha} - \frac{e^{-h_1\rho}}{\rho-\alpha}\frac{1}{s+\rho}\right] - e^{-(T_0+h_1)\alpha} \times \left[\frac{1}{\rho-\alpha}\frac{1}{s+\alpha} - \frac{1}{\rho-\alpha}\frac{1}{s+\rho}\right]\right\} \quad (28)$$

Therefore

$$P_{T_0+h_1}{}^{\infty}[C_0(s)] = e^{-(T_0+h_1)s} \times \left[\frac{e^{-T_0\alpha}}{\alpha-\rho}(e^{-\rho h_1} - e^{-\alpha h_1})\frac{1}{s+\rho}\right] \quad (29)$$

which is identical to equation 24.

The steps shown by equations 27 and 28 may be eliminated by the use of Table I. In this case relation 5 of Table I, together with equation 14, can be used to write down equation 29 directly from equation 26.

The generalization, and extension of these results, to samplers operating for a finite or infinite number of pulses is now a trivial matter. Since it is assumed that $KG(s)$ represents a linear system, the laws of superposition may be applied. The sampler output may be written as the sum of individual pulses as follows:

$$E_p(s) = \sum_{n=0}^{\infty} P_{T_n}{}^{T_n+h_{n+1}}[E(s)] \quad (30)$$

The total output therefore becomes:

$$C(s) = \sum_{n=0}^{\infty} C_n(s) = KF(s)\left\{\sum_{n=0}^{\infty} \times P_{T_n}{}^{T_n+h_{n+1}}[E(s)]\right\} \quad (31)$$

When the sampler is periodic, with period T and uniform pulse width h, this summation can be expressed in a closed form as shown in references 5 and 6.

Physical reasoning, together with equation 30, indicate that a sampler operating for a finite or infinite number of pulses may be replaced by a finite or infinite number of samplers operating in parallel where each of these operates only once and has a single output pulse. That is, the first sampler closes at T_0 and opens at T_0+h_1 remaining open thereafter. The second closing at T_1, opening at T_1+h_2, remaining open thereafter, and so on. Assuming that the sampler shown in Fig. 1 operates for an infinite number of pulses, then in view of the above and equation 31 the system of Fig. 1 may be replaced by an equivalent system shown in Fig. 3, where each sampler has a single output pulse, stretching from T_n to T_n+h_{n+1}. These pulses are fed to a system with transfer function $KG(s)$ in order to obtain $C_n(s)$, their corresponding component of output. These components are then fed to a summator, the output of which is $C(s)$ the total output.

Extension to Feedback Systems

Fig. 4(A) shows a finite pulsed feedback system. The object in analyzing such a system is to determine the output $c(t)$ for a specified input $r(t)$. Two general assumptions will be made regarding all feedback systems considered.

1. At $t=0$ all initial conditions are zero.
2. Systems have a continuous step response. That is, the denominator of $KG(s)$ is one or more degrees higher than its numerator.

Assumption 1 is made so that the system under consideration can be represented by a transfer function $KG(s)$. Assumption 2 eliminates the discontinuities which would otherwise occur in the output due to the step characteristics of the input pulses. This means the output at the end of one period is the same as beginning of the next, a characteristic which is used in the following analysis. These assumptions may be modified where necessary. However, it is found that in general most systems satisfy the above requirements.

In the following, the feedback system shown in Fig. 4(A) is studied. The principles applied to its analysis can be used with minor modifications for analysis of more general configurations.

Referring to Fig. 4(A), let the sampler, not necessarily periodic, close at specified instants T_n, remain closed for duration h_{n+1} seconds, and then open at instants T_n+h_{n+1}. The pulse widths h_{n+1} and the sampling instants T_n may be dependent on any desired quantity. This does not affect the computations so long as the instants at which they occur are known or can be solved for.

Fig. 4(A) can be replaced by its equivalent system shown in Fig. 4(B). This consists of infinite number of samplers operating in parallel, each sampler closing and opening only once, the nth sampler closing at instant T_n, opening at T_n+h_{n+1} seconds, and remaining open thereafter. Consider the total output $C(s)$ during the interval $0 \le t \le T_1$ (the limit points $t=0$ and $t=T_1$ can be included because of assumption 2). With zero initial conditions the only component contributing to the total output in this interval is $C_0(s)$. This is because the other samplers have not yet closed and hence cannot contribute to the total output. Therefore,

$$C(s) = C_0(s) \text{ for } 0 \le t \le T_1 \quad (32)$$

and the system of Fig. 4(B) can be represented by that shown in Fig. 5 for this duration.

Consider now the first input pulse $E_{0p}(s)$ and its component of output $C_0(s)$. Referring to Fig. 5, it is evident that while the sampler is closed,

$$e_{0p}(t) = e_0(t) \quad 0 < t < h_1 \quad (33)$$

During this interval the output and hence the error are also indentical to the continuous system.

Therefore

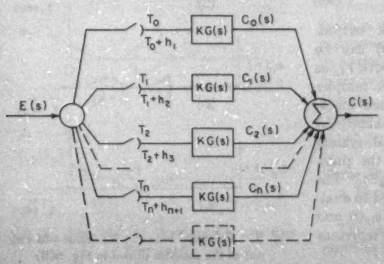


Fig. 3 (left). Equivalent system of Fig. 1

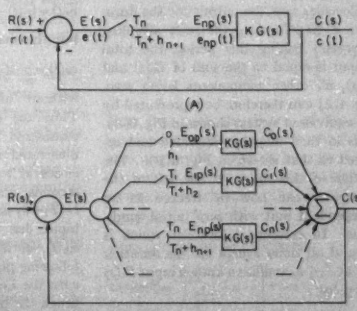


Fig. 4 (right). Finite pulsed feedback system and its equivalent

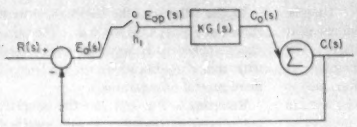
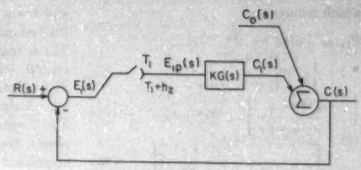


Fig. 5 (left). Equivalent system of Fig. 4 for the interval $0 \leq t \leq T_1$



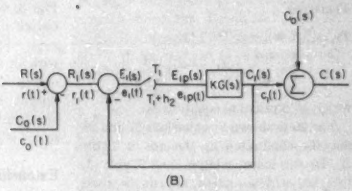


Fig. 6 (right). Equivalent system of Fig. 4 for the interval $T_1 \leq t \leq T_2$

$$E_{0}(s) = \frac{R(s)}{1+KG(s)} \quad 0<t<h_1 \tag{34}$$

At $t=h_1$ the sampler is opened and hence the input is zero for all t greater than h_1. Thus the following equations can be written:

$$e_{0p}(t) = \begin{cases} e_0(t) \text{ for } 0<t<h_1 \\ 0 \text{ for } t>h_1 \end{cases} \tag{35}$$

which may be expressed as:

$$e_{0p}(t) = e_0(t)u_0^{h_1}(t) \text{ for all } t>0 \tag{36}$$

The P-notation can now be used to write the Laplace transform of equation 36 as:

$$E_{0p}(s) = P_0^{h_1}\left[\frac{R(s)}{1+KG(s)}\right] \quad \text{for all } t>0 \tag{37}$$

The inverse of equation 37 satisfies all the conditions imposed on $e_{0p}(t)$ by equations 33 and 35. Therefore $C_0(s)$, the component of output due to the first pulse, becomes:

$$C_0(s) = E_{0p}(s)KG(s) = P_0^{h_1}\left[\frac{R(s)}{1+KG(s)}\right]KG(s) \tag{38}$$

Equation 38 describes $C_0(s)$ and hence its inverse $c_0(t)$ for all values of time. In view of this and relation 32, the total output for the duration $0 \leq t \leq T_1$ can be written as:

$$C(s) = P_0^{h_1}\left[\frac{R(s)}{1+KG(s)}\right]KG(s) \text{ for } 0 \leq t \leq T_1 \tag{39}$$

Since both $R(s)$ and $KG(s)$ are known, the output $c(t)$ may be evaluated from the inverse of equation 39 for this interval.

Consider now the output for the duration $T_1 \leq t \leq T_2$, i.e., the next sampling interval. During this period the total output is equal to the sum of $C_0(s)$ and $C_1(s)$, all other components being zero. Fig. 4(A) can therefore be represented by the equivalent system shown in Fig. 6(A). This in turn may be simplified and reduced to that shown in Fig. 6(B). The section of this figure to the left of the summator can now be treated as an independent unit with zero initial conditions. In contains a sampler which is open at all times except for the duration $T_1<t<T_1+h_2$ and has a known input $R_1(s)$ given by:

$$R_1(s) = [R(s) - C_0(s)] \tag{40}$$

$C_1(s)$, the output of this system, may be evaluated in the following manner. Suppose that the input $r_1(t)$ is the time function shown in Fig. 7(A). The value of this for the duration $0<t<T_1$ has no effect whatsoever on the output $c_1(t)$, because the sampler is open during this interval. It may therefore be assumed to be zero and conveniently replaced by a new function $r_{11}(t)$ as shown in Fig. 7(B).

Where

$$r_{11}(t) = \begin{cases} 0 \text{ for } 0<t<T_1 \\ r_1(t) \text{ for } t>T_1 \end{cases} \tag{41}$$

equation 41 can also be written as:

$$r_{11}(t) = r_1(t)\, u_{T_1}^{\infty}(t) \tag{42}$$

Using the P-notation, the Laplace transform of this equation becomes:

$$R_{11}(s) = P_{T_1}^{\infty}[R_1(s)] = P_{T_1}^{\infty}[R(s) - C_0(s)] \tag{43}$$

With zero initial conditions the output $c_1(t)$ will also be zero for the interval $0<t<T_1$. This is illustrated graphically in Fig. 7(C). The error is given by:

$$e_1(t) = [r_1(t) - c_1(t)] \tag{44}$$

but this may be replaced by:

$$e_{11}(t) = [r_{11}(t) - c_1(t)] \tag{45}$$

without affecting the actual output. Thus the effective error $e_{11}(t)$ can be considered to be zero for $0<t<T_1$ as illustrated in Fig. 7(D). The sampler closes at $t=T_1$, remaining closed for h_2 seconds, and is opened permanently at $t=T_1+h_2$ seconds. The actual system input due to this closure is the pulse $e_{1p}(t)$, as illustrated in Fig. 7(E). The following procedures can be used to evaluate the Laplace transform of $e_{1p}(t)$ and hence the output $c_1(t)$. Let t_1 represent a new time axis such that

$$t_1 = (t - T_1) \tag{46}$$

$r_{11}(t)$, $c_1(t)$, $e_{11}(t)$, and $e_{1p}(t)$ may now be expressed in terms of t_1 as:

$$\begin{aligned} &\text{(A)}\ r_{11}(t) = d(-T_1)r_{11}(t_1) \\ &\text{(B)}\ c_1(t) = d(-T_1)c_1(t_1) \\ &\text{(C)}\ e_{11}(t) = d(-T_1)e_{11}(t_1) \\ &\text{(D)}\ e_{1p}(t) = d(-T_1)e_{1p}(t_1) \end{aligned} \tag{47}$$

where the new functions $r_{11}(t_1)$, $c_1(t_1)$, $e_{11}(t_1)$, and $e_{1p}(t_1)$ are as shown in Figs. 8(A), 8(B), 8(C), and 8(D) respectively, and $d(-T_1)$ is the notation used to express a delay of T_1 seconds along the time axis.

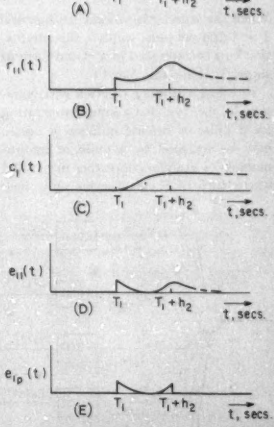


Fig. 7. Relation between the input and output of the sampler shown in Fig. 6(B)

The Laplace transforms of equations 47 become:

$$\begin{array}{ll}(A) & R_{11}(s)=e^{-T_1 s}\mathcal{L}[r_{11}(t_1)]\\ (B) & C_1(s)=e^{-T_1 s}\mathcal{L}[c_1(t_1)]\\ (C) & E_{11}(s)=e^{-T_1 s}\mathcal{L}[c_{11}(t_1)]\\ (D) & E_{1p}(s)=e^{-T_1 s}\mathcal{L}[e_{1p}(t_1)]\end{array} \qquad (48)$$

The situation is now identical to that of the first pulse with respect to the time axis t_1. That is, an input $r_{11}(t_1)$ is sampled for duration $0<t_1<h_2$ and fed to a system with zero initial conditions. Following the same procedure as before the system equations with respect to time axis t_1 can be written as:

$$\mathcal{L}[c_{11}(t_1)]=\frac{\mathcal{L}[r_{11}(t_1)]}{1+KG(s)} \qquad 0<t_1<h_2 \qquad (49)$$

and therefore,

$$\mathcal{L}[e_{1p}(t_1)]=P_0^{h_2}\left[\frac{\mathcal{L}[r_{11}(t_1)]}{1+KG(s)}\right] \text{ for all } t_1>0 \qquad (50)$$

The Laplace transform of the output may now be written as:

$$\mathcal{L}[c_1(t_1)]=P_0^{h_2}\left[\frac{\mathcal{L}[r_{11}(t_1)]}{1+KG(s)}\right]KG(s) \qquad (51)$$

In order to obtain the total output it is necessary to express the above equation in terms of the original time axis t. This can be done by means of equation 48. Using relations 48(A), 48(B), equation 51 can be written as:

$$e^{T_1 s}C_1(s)=P_0^{h_2}\left[\frac{e^{T_1 s}R_{11}(s)}{1+KG(s)}\right]KG(s) \qquad (52)$$

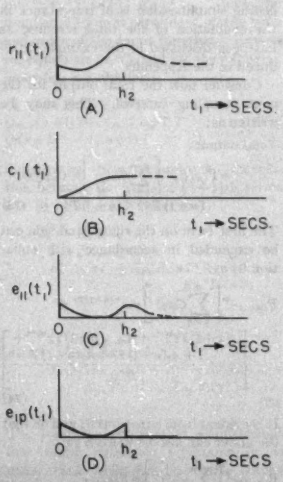
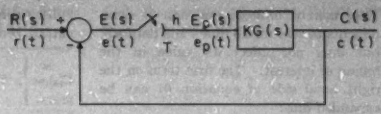


Fig. 8. Showing Fig. 7 referred to the new time axis t_1

Fig. 9. Error pulsed feedback system with a periodic sampler



Substituting equation 43 for $R_{11}(s)$ in the above, it becomes:

$$C_1(s)=e^{-T_1 s}P_0^{h_2}\left[\frac{e^{T_1 s}P_{T_1}^{\infty}[R(s)-C_0(s)]}{1+KG(s)}\right]\times KG(s) \qquad (53)$$

Equation 53 expresses $C_1(s)$ and hence its inverse $c_1(t)$ for all values of time t greater than zero. The total output for the duration $T_1\leq t\leq T_2$ can now be written as:

$$C(s)=C_0(s)+C_1(s) \quad T_1\leq t\leq T_2 \qquad (54)$$

which may be determinated from equations 38 and 53. Proceeding in this manner the total output may be evaluated at any instant of time. It is not difficult to generalize and show that $C_n(s)$, the component of output due to the $(n+1)$th pulse, is:

$$C_n(s)=e^{-T_n s}P_0^{h_{n+1}}\times\left[\frac{e^{T_n s}P_{T_n}^{\infty}[R(s)-C_0(s)-\dfrac{C_1(s)-\ldots-C_{n-1}(s)]}{1+KG(s)}}\right]KG(s) \qquad (55)$$

from which the total output for the duration $T_n\leq t\leq T_{n+1}$ can be obtained using the relation:

$$C(s)=\sum_{k=0}^{n}C_k(s) \text{ for } T_n\leq t\leq_{n+1} \qquad (56)$$

In numerical calculations, equation 55 for $C_n(s)$ is simplified by the fact that at $t=T_n$ all the components $C_0(s)$, $C_1(s)$, ... $C_{n-1}(s)$ must have the same poles as $KG(s)$. This, as was pointed out previously in relation 16(C) is so because after the sampler is opened there is no feedback and the rise or decay of the system is dependent entirely on the poles of $KG(s)$. The sum of these components can therefore be expressed as a polynomial whose denominator is identical to that of $KG(s)$ and whose numerator is dependent on the value of T_n.

The foregoing illustrates how the superposition principle, together with some properties of the p-transform, can be used in evaluating the output of a closed-loop finite pulsed system with a nonperiodic sampler. The step-by-step analysis is necessary whenever the sampler operation is not governed by a repetitive pattern.

Error Pulsed Feedback System With a Periodic Sampler

The general method developed in the previous section is now applied to analysis of an error pulsed feedback system with a periodic sampler. The periodic operation of the sampler leads to some simplification, but it also raises the question of stability. The stability criteria for such systems can be expressed in terms of a set of constant coefficients. These coefficients are referred to as the characteristic coefficients, because they determine the system characteristic equation. The roots of this equation then determine the system stability.

Fig. 9 shows an error pulsed system with a periodic sampler having a period T and uniform pulse width h seconds. The transfer function can in general be expressed as the ratio of two polynomials in s. For systems being considered this may be written as:

$$KG(s)=\frac{a_{m-1}s^{m-1}+a_{m-2}s^{m-2}+\ldots\ldots+a_1 s+a_0}{s^m+b_{m-1}s^{m-1}+\ldots\ldots+b_1 s+b_0}=\frac{N(s)}{D(s)} \qquad (57)$$

where m is the order of the system, and the numerator is at least one degree lower than the denominator in accordance with assumption 2.

Consider the outputs for the $(n+1)$th sampling period, i.e., for duration $nT\leq t\leq (n+1)T$. The total output for this interval can be written as:

$$C(s)=\sum_{k=0}^{n}C_k(s) \text{ for } nT\leq t\leq (n+1)T \qquad (58)$$

where $n=0, 1, 2, \ldots\ldots$ all positive integers. $C_n(s)$, the component of output due to the pulse stretching from $nT<t<nT+h$ can be obtained from equation 56 and is:

$$C_n(s)=e^{-nTs}P_0^{h}\times\left[\frac{e^{nTs}P_{nT}^{\infty}[R(s)-C_0(s)-\dfrac{C_1(s)-\ldots-C_{n-1}(s)]}{1+KG(s)}}\right]KG(s) \qquad (59)$$

Note that $c_n(t)$, the inverse of this, is a continuous function of time starting with zero amplitude. This is because $KG(s)$ is assumed to have a continuous step response. Therefore,

$$c_n(nT)=0 \qquad (60)$$

Equation 58 can be written as:

$$C(s)=P_{nT}^{\infty}\left[\sum_{k=0}^{n-1}C_k(s)\right]+C_n(s) \qquad nT<t\leq(n+1)T \qquad (61)$$

This merely reduces the first n components to zero for all time t less than nT, but does not affect its value in the region of interest. The first term on the right hand side of equation 61 can be expanded into:

$$P_{nT}^{*}\left[\sum_{k=0}^{n-1} C_k(s)\right] = \{P_{nT}^{*}[C_0(s)] + P_{nT}^{*}[C_1(s)] + \ldots\ldots + P_{nT}^{*}[C_{n-1}(s)]\} \tag{62}$$

Some general characteristics of this term can now be determined from its individual components. For example, consider the first component

$$P_{nT}^{*}[C_0(s)]$$

The general properties of this function may be summarized as follows:

1. Its inverse, $c_0(t)$, is zero for all $t<nT$. This is by virtue of the p-transformation, and assures the presence of e^{-nTs} as a multiplier in the s-plane.

2. At $t=nT$, $c_0(nT)$ is finite in amplitude for all finite n.

3. For $t>h$, or $n\geq 1$, its denominator is identical to that of $KG(s)$. The reason for this has been pointed out in equation 16(C).

From these three characteristics, the general equation for this term can be written as:

$$P_{nT}^{*}[C_0(s)] = e^{-nTs}\times \left[\frac{x_{m-1}{}^{0}s^{m-1}+x_{m-2}{}^{0}s^{m-2}+ \ldots\ldots +x_1{}^{0}s+x_0{}^{0}}{D(s)}\right] \tag{63}$$

where the x's are dependent on the system constants as well as the value of nT, and $D(s)$ is the denominator of $KG(s)$ as in equation 57. Note that the numerator of equation 63 is at least one degree less than its denominator. This is because $c_0(t)$ is finite at $t=nT$. It is not difficult to verify the validity of this equation for any specified system. The above reasoning can be extended to all other terms on the right hand side of equation 62 to show that they all have e^{-nTs} as multiplier and $D(s)$ as denominator. Therefore, they can be placed on a common denominator and written as:

$$P_{nT}^{*}\left[\sum_{k=0}^{n-1} C_k(s)\right] = e^{-nTs}\times \left[\frac{A_{m-1}(nT)s^{m-1}+A_{m-2}(nT)s^{m-2}+ \ldots +A_1(nT)s+A_0(nT)}{D(s)}\right] \tag{64}$$

where

$$\begin{aligned} A_{m-1}(nT) &= x_{m-1}{}^{0}+x_{m-1}{}^{1}+\ldots\ldots+x_{m-1}{}^{n-1} \\ &\vdots \\ A_0(nT) &= x_0{}^{0}+x_0{}^{1}+\ldots\ldots+x_0{}^{n-1} \end{aligned} \tag{65}$$

Substituting equation 64 in 61, the total output becomes:

$$C(s) = \left\{ e^{-nTs}\times \left[\frac{A_{m-1}(nT)s^{m-1}+A_{m-2}(nT)\times s^{m-2}+\ldots\ldots+A_0(nT)}{D(s)}\right] + C_n(s)\right\} \quad nT<t\leq(n+1)T \tag{66}$$

The initial value theorem can be applied to this delayed function to show that:

$$A_{m-1}(nT) = c(nT) \tag{67}$$

This is because $c_n(t)$, the component due to the last term, is zero for all $0<t<nT$ and at $t=nT$. A purely algebraic process can also be used to verify equation 67. In a similar way it can be shown that the other coefficients are related to the higher order derivatives at $t=nT$. However, these are not of any interest in the present analysis but may be employed advantageously in a synthesis problem. The total output can now be written from equations 59, 64, 66, and 67 as:

$$C(s) = e^{-nTs}\left\{ \frac{c(nT)s^{m-1}+A_{m-2}\times s^{m-2}+(nT)s^{m-2}+\ldots\ldots+A_1(nT)s+A_0(nT)}{D(s)} + P_0{}^{h}\left[\frac{e^{nTs}P_{nT}^{*}[R(s)]-\dfrac{c(nT)s^{m-1}+A_{m-2}(nT)\times s^{m-2}+\ldots\ldots+A_1(nT)s+\dfrac{A_0(nT)}{D(s)}}{1+KG(s)}}\right] KG(s)\right\} \tag{68}$$

Equation 68 and its inverse describe the continuous output for duration $nT<t\leq(n+1)T$. It is possible to simplify this equation in the following manner. $C(s)$ can be divided into two components, $C_h(s)$ valid for $nT<t\leq nT+h$, i.e., while sampler is closed, and $C_T(s)$ valid for $nT+h<t\leq(n+1)T$, i.e., while sampler is open. The symbol $P_0{}^{h}$ can be disregarded for $nT<t\leq nT+h$, because it signifies multiplication by unity during this interval. Therefore, the first component of output can be written as:

$$C_h(s) = e^{-nTs}\times \left\{\frac{c(nT)s^{m-1}+A_{m-2}(nT)s^{m-2}+\ldots\ldots+A_0(nT)}{D(s)}\times \left[1-\frac{KG(s)}{1+KG(s)}\right]+e^{nTs}P_{nT}^{*}[R(s)]\times \frac{KG(s)}{1+KG(s)}\right\} \quad nT<t\leq nT+h \tag{69}$$

which reduces to:

$$C_h(s) = e^{-nTs}\times \left\{\frac{c(nT)s^{m-1}+A_{m-2}(nT)s^{m-2}+\ldots\ldots+A_0(nT)}{N(s)+D(s)}+ e^{nTs}P_{nT}^{*}[R(s)]\frac{N(s)}{N(s)+D(s)}\right\} \quad nT<t\leq nT+h \tag{70}$$

where $N(s)$ and $D(s)$ are the numerator and denominator of $KG(s)$ as given by equation 57. It is seen that during this interval the output has the same poles as $1/1+KG(s)$. The second component can be written as:

$$C_T(s) = P_{(nT+h)}^{*}[C(s)] \quad nT+h<t\leq(n+1)T \tag{71}$$

$C_T(s)$ must have the same poles as $KG(s)$ because during this interval the sampler is open. In general, $C_T(s)$ can be shown to be in the form of (see Appendix):

$$C_T(s) = e^{-(nT+h)s}\left[\sum_{k=1}^{m} L_c{}^{k}c(nT)+ L_{m-2}{}^{k}A_{m-2}(nT)+\ldots\ldots+ L_0{}^{k}A_0(nT)+L_{R1}{}^{k}R_1(nT)+ L_{R2}{}^{k}R_2(nT)+\ldots\ldots\right]\frac{e^{-\rho_k h}}{s+\rho_k} \quad nT+h<t\leq(n+1)T \tag{72}$$

where $L_c{}^{k}\ldots L_0{}^{k}$, $L_R{}^{k}$ are constants, $R_1(nT)$, $R_2(nT)\ldots$ are functions dependent on the type of input, and ρ_1, $\rho_2\ldots\rho_m$ are the simple poles $KG(s)$. $C_T(s)$ has a slightly different form if $KG(s)$ has multiple poles. The value of the preceding simplification is of importance in the evaluation of the total response as has been discussed in the example introduced in the Appendix.

Consider now the total output for the next sampling interval. This may be written as:

Total output

$$= P_{(n+1)T}^{*}\left[\sum_{k=0}^{n} C_k(s)\right] + C_{n+1}(s) \quad (n+1)T<t\leq(n+2)T \tag{73}$$

The first term on the right hand side can be expanded in accordance with equation 64 as:

$$P_{(n+1)T}^{*}\left[\sum_{k=0}^{n} C_k(s)\right] = e^{-(n+1)Ts}\times \left[\frac{c(n+1)Ts^{m-1}+A_{m-2}(n+1)Ts^{m-2}+ \ldots +A_1(n+1)Ts+A_0(n+1)T}{D(s)}\right] \tag{74}$$

It is evident from equations 61 and 68 that the above can also be written as:

$$P_{(n+1)T}^{*}\left[\sum_{k=0}^{n} C_k(s)\right] = P_{(n+1)T}^{*} \quad \text{[equation 68]} \tag{75}$$

Table II. Characteristic Coefficients and the Roots z_1 and z_2 for Various Values of the Pulse Width h for Second-Order System Considered with T=1 Second, $p_2=5$, and K=6

h, Seconds	K_c^0	K_c^1	K_c^2	K_c^3	z_1	z_2
0.2	−0.1375	0.1815	−0.7290	0.9133	0.7671	0.0088
0.4	−0.1687	0.1491	−0.8888	0.7456	0.5650	0.0119
0.6	−0.1554	0.1175	−0.8153	0.5730	0.4008	0.0158
0.8	−0.1328	0.0945	−0.6970	0.4342	0.2662	0.0253
T=1	−0.1213	0.0855	−0.5133	0.3064	0.1353	0.0498

For $t \geq (n+1)T$ the right hand side of equation 75 can be shown to be:

$$P_{(n+1)T^m}\ [\text{equation 68}] = \varepsilon^{-(n+1)Tz} \times \left[\frac{Q_{m-1}(n+1)Tz^{m-1} + Q_{m-2} \times (n+1)Tz^{m-2} + \ldots + Q_0(n+1)T}{D(s)} \right] \quad (76)$$

where the Q's are functions of $c(nT)$, $A_{m-2}(nT) \ldots . A_0(nT)$, as well as the pulse width h and the constants of transfer function. These can in general be written as (see Appendix):

$$\begin{aligned} Q_{m-1}(n+1)T &= K_c{}^c c(nT) + K_{m-2}{}^c A_{m-2}(nT) + \ldots . + K_0{}^c A_0(nT) + K_{R_1}{}^c R_1(nT) + K_{R_2}{}^c R_2(nT) + \ldots \\ &\vdots \\ Q_0(n+1)T &= K_c{}^0 c(nT) + K_{m-2}{}^0 A_{m-2}(nT) + \ldots . + K_0{}^0 A_0(nT) + K_{R_1}{}^0 R_1(nT) + K_{R_2}{}^0 R_2(nT) + \ldots . \end{aligned} \quad (77)$$

where the K_c, K_{m-2}, ... K_1, K_0 are called the "characteristic coefficients" and are constants dependent only on the system parameters. The constants K_R, the functions $R_1(nT)$, $R_2(nT) \ldots .$ and the number of them is dependent on the type and order of the input function. Equating equations 74, 75, and 76, the following identities are obtained:

$$\begin{aligned} Q_{m-1}(n+1)T &= c(n+1)T \\ Q_{m-2}(n+1)T &= A_{m-2}(n+1)T \\ Q_0(n+1)T &= A_0(n+1)T \end{aligned} \quad (78)$$

Substituting these in equations 77, a relation between the nth and $(n+1)$th terms can be obtained. These are:

$$\begin{aligned} c(n+1)T &= K_c{}^c c(nT) + K_{m-2}{}^c A_{m-2}(nT) + \ldots . + K_0{}^c A_0(nT) + K_{R_1}{}^c R_1(nT) + K_{R_2}{}^c R_2(nT) \ldots . \\ &\vdots \\ A_0(n+1)T &= K_c{}^0 c(nT) + K_{m-2}{}^0 A_{m-2}(nT) + \ldots . + K_0{}^0 A_0(nT) + K_{R_1}{}^0 R_1(nT) + K_{R_2}{}^0 R_2(nT) + \ldots . \end{aligned} \quad (79)$$

The foregoing provides m simultaneous first order difference equations in m unknowns. They can be readily solved by the z-transform method in the following manner. The z-transform of $c(n+1)T$ may be written as:

$$Z[c(n+1)T] = z[C(z) - c(0)] \quad (80)$$

where

$$C(z) = Z[c(nT)] \quad (81)$$

and

$$z = \varepsilon^{sT} \quad (82)$$

Similarly,

$$\begin{aligned} Z[A_{m-2}(n+1)T] &= z[A_{m-2}(z) - A_{m-2}(0)] \\ &\vdots \\ Z[A_0(n+1)T] &= z[A_0(z) - A_0(0)] \end{aligned} \quad (83)$$

With zero initial conditions all the terms $c(0)$, $A_{m-2}(0) \ldots . A_0(0)$ are zero. The z-transform of equations 79 can now be written with the aid of the preceding equations as:

$$\begin{aligned} (z - K_c{}^c) C(z) - K_{m-2}{}^c A_{m-2}(z) - \ldots . - K_0{}^c A_0(z) &= K_{R_1}{}^c R_1(z) + K_{R_2}{}^c R_2(z) + \ldots . \\ -K_c{}^{m-2} C(z) + (z - K_{m-2}{}^{m-2}) A_{m-2}(z) - \ldots . - K_0{}^{m-2} A_0(z) &= K_{R_1}{}^{m-2} R_1(z) + K_{R_2}{}^{m-2} R_2(z) + \ldots . \\ &\vdots \\ -K_c{}^0 C(z) - K_{m-2}{}^0 A_{m-2}(z) - \ldots . - (z - K_0{}^0) A_0(z) &= K_{R_1}{}^0 R_1(z) + K_{R_2}{}^0 R_2(z) + \ldots . \end{aligned} \quad (84)$$

where the only unknowns are $C(z)$, $A_{m-2}(z) \ldots . A_0(z)$. Equations 84 can be solved by the determinant method: Let

$$\Delta(z) = \begin{vmatrix} (z - K_c^2) & -K_{m-2}{}^c \ldots . & -K_0{}^c \\ -K_c{}^{m-2}(z - K_{m-2}{}^{m-2}) \ldots . & -K_c{}^{m-2} \\ \vdots & \\ -K_c{}^0 & -K_{m-2}{}^0 \ldots . & (z - K_0{}^0) \end{vmatrix} \quad (85)$$

$C(z)$, the z-transform of the output can then be written as:

$$C(z) = \frac{\begin{vmatrix} K_{R_1}{}^c R_1(z) + K_{R_2}{}^c R_2(z) + \ldots . - K_{m-2}{}^c - K_{m-2}{}^c \ldots . - K_0{}^c \\ \vdots \\ K_{R_1}{}^0 R_1(z) + K_{R_2}{}^0 R_2(z) + \ldots . - K_{m-2}{}^0 - K_{m-2}{}^0 \ldots . (z - K_0{}^0) \end{vmatrix}}{\Delta(z)} \quad (86)$$

The inverse of equation 86 gives the values of the output at instants $t = nT$. $\Delta(z)$ is the characteristic equation whose roots determine the stability. For a stable system these roots must be within the unit circle. For the mth order system under consideration $\Delta(z)$ can be represented by an mth order polynomial in z as:

$$\Delta(z) = z^m + B_{m-1} z^{m-1} + \ldots . + B_1 z + B_0 \quad (87)$$

The constant coefficients $B_{m-1} \ldots . B_0$ can be expressed directly in terms of the characteristic coefficients as is seen from relation 85. Note that they are independent of the input coefficients $K_{R_1} \ldots .$ etc. It is possible to develop the stability criteria in terms of the relation between the B coefficients.[11] For example, it can be shown that for a second-order system:

$$\Delta(z) = z^2 + B_1 z + B_0 \quad (88)$$

and for stability[11]

$$\left.\begin{aligned} \frac{1 - B_0}{1 + B_1 + B_0} &> 0 \\ \frac{1 - B_1 + B_0}{1 + B_1 + B_0} &> 0 \end{aligned}\right\} \quad (89)$$

The relations between the B's, the characteristic coefficients, and the system parameters, together with the stability conditions for a general second-order transfer function are derived in the Appendix. Numerical substitutions in those general equations will immediately determine the stability of a given system. The same procedure can be used to derive the equations for a third and higher order configurations.

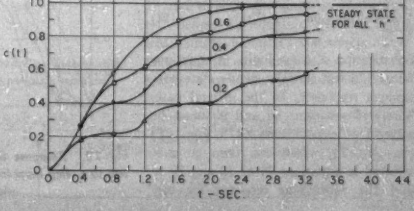


Fig. 10. Response of the second-order system to a unit step input for various values of pulse width h

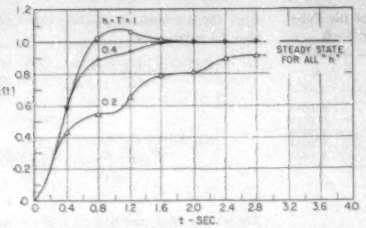


Fig. 11. Response of the second-order system to a unit step input when K=15.25 and the closed loop poles are complex

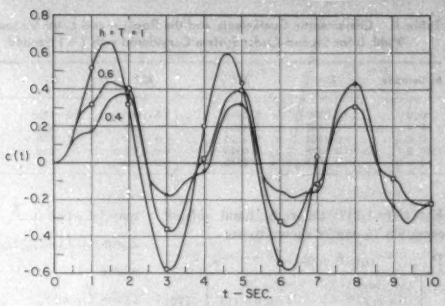


Fig. 13. Response of the second order system to the sinusoidal input $R(s)=2/s^2+4$

If only the stability and the output at $t=nT$ is desired, then equation 86 suffices to describe the system. In general it is advisable to determine the continuous output in order to examine its ripples. In this case it becomes necessary to solve for $A_{m-2}(z),\ldots.A_0(z)$ as follows:

$$A_{m-2}(z)=\frac{\begin{vmatrix} (z-K_c{}^c)K_{R_1}{}^cR_1(z)+K_{R_2}{}^cR_2(z)\times\ldots.-K_{m-2}{}^c\ldots.-K_0{}^c \\ -K_c{}^{m-2}K_{R_1}{}^{m-2}R_1(z)+K_{R_2}{}^{m-2}\times R_2(z)\ldots.-K_{m-2}{}^{m-2}\ldots. \\ \qquad\qquad -K_0{}^{m-2} \\ \vdots \\ -K_c{}^0K_{R_1}{}^0R_1(z)+K_{R_2}{}^0R_2(z)\ldots. \\ \qquad\qquad -K_{m-2}{}^0\ldots.(z-K_0{}^0) \end{vmatrix}}{\Delta(z)}$$

$$A_0(z)=\frac{\begin{vmatrix} z-K_c{}^c-K_{m-2}{}^c-K_{m-1}{}^c\ldots. \\ \qquad -K_1{}^cK_{R_c}{}^cR_1(z)+K_{R_2}R_2(z) \\ -K_c{}^{m-2}(z-K_{m-2}{}^{m-2})\ldots.\ldots. \\ \qquad K_{R_1}{}^{m-2}R_1(z)+K_{R_2}R_2(z) \\ \vdots \\ -K_c{}^0-K_{m-2}{}^0\ldots.K_{R_1}{}^0R_1(z)+ \\ \qquad K_{R_2}{}^{m-2}R_2(z) \end{vmatrix}}{\Delta(z)} \tag{90}$$

The steady-state ripple can usually be obtained directly from the equations for $C(z)$, $A_{m-2}(z)\ldots.A_0(z)$. Assuming the limits exist, the final value theorem[2] can be applied to these equations to determine the steady-state values of the corresponding coefficients. For example,

$$\lim_{n\to\infty} c(nT)=\lim_{z\to 1}\frac{z}{z-1}C(z) \tag{91}$$

Substitution of these limits in equations 70 and 72 will give the steady state continuous response. In the general case, however, equations 86 and 90 must be solved for the coefficients $c(nT)$, $A_{m-2}(nT)$ $\ldots.A_0(nT)$, which can then be substituted in equations 68 or 70 and 72 to find the continuous response for any desired interval $nT\leq t\leq(n+1)T$. Note that all these coefficients have the same characteristic equation $\Delta(z)$. Thus having once determined the roots of $\Delta(z)$, they can be used to evaluate all the other coefficients. This reduces the labor by a considerable amount.

So far the analysis has been carried out in a general form. This must now be applied to a specific example in order to illustrate some of the details. Referring to Fig. 9, let

$$R(s)=\frac{1}{s} \tag{92}$$

and

$$KG(s)=\frac{K}{s(s+\rho_2)} \tag{93}$$

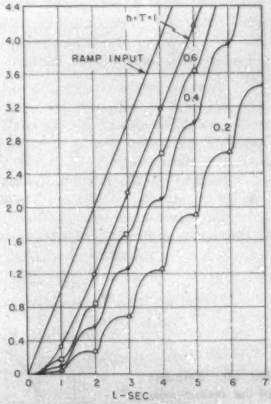


Fig. 12. Response of the second order system to a ramp input for various values of pulse width h

Therefore,

$$\frac{1}{1+KG(s)}=\frac{s(s+\rho_2)}{s^2+\rho_2 s+K}=\frac{s(s+\rho_2)}{(s+\alpha_1)(s+\alpha_2)} \tag{94}$$

where

$$\alpha_1=\frac{\rho_2}{2}+\beta$$

$$\alpha_2=\frac{\rho_2}{2}-\beta \tag{95}$$

and

$$\beta=\sqrt{\left(\frac{\rho_2}{2}\right)^2-K}$$

The equations for this system can be obtained directly from those given in the Appendix by substitution of:

$$\begin{aligned} a_1&=0\\ a_0&=K\\ \rho_1&=0\\ M&=1 \end{aligned} \tag{96}$$

$C_h(s)$, the output for $nT<t\leq nT+h$, is obtained from equation 141 and is:

$$C_h(s)=e^{-nTs}\times\left[\frac{c(nT)s+A_0(nT)}{(s+\alpha_1)(s+\alpha_2)}+\frac{K}{s(s+\alpha_1)(s+\alpha_2)}\right] \qquad nT<t\leq(nT+h) \tag{97}$$

Similarly, the output for $nT+h<t\leq(n+1)T$ can be determined from equation 142 and is:

$$C_T(s)=e^{-(nT+h)s}\left\{[L_c{}^1c(nT)+L_0{}^1A_0(nT)+L_R{}^1]\frac{1}{s}+[L_c{}^2c(nT)+L_0{}^2A_0(nT)+L_R{}^2]\frac{e^{-\rho_2 h}}{s+\rho_2}\right\} \qquad nT+h<t\leq(n+1)T \tag{98}$$

where the L coefficients are given by equations 132. Relations 139 and 140 can be used to write $C(z)$ and $A_0(z)$ as:

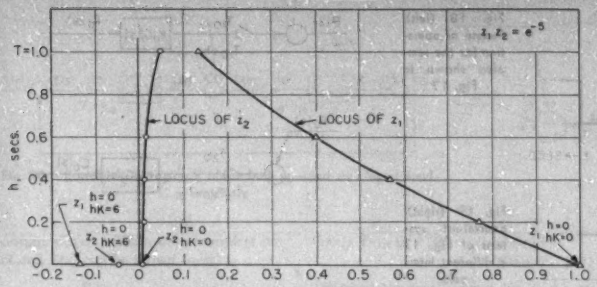


Fig. 14. Plot of z_1 and z_2 as a function of the pulse width h

$$C(z)=\frac{z[K_R{}^0z-(K_R{}^0K_0{}^0-K_c{}^cK_R{}^0)]}{(z-1)(z-z_1)(z-z_2)} \quad (99)$$

and

$$A_0(z)=\frac{z[K_R{}^0z-(K_c{}^cK_R{}^0-K_c{}^0K_R{}^c)]}{(z-1)(z-z_1)(z-z_2)} \quad (100)$$

where the coefficients $K_R{}^0$, $K_c{}^c$ ..... etc. are given by equations 134 of the Appendix and z_1 and z_2 are the roots of the characteristic equations as given by:

$$\Delta(z)=z^2+B_1z+B_0=(z-z_1)(z-z_2) \quad (101)$$

In this example it can be shown that:

$$B_1=-e^{-\frac{\rho_2 h}{2}}\left\{[1+e^{-\rho_2(T-h)}]\cosh\beta h+\left(\frac{\rho_2}{2}-\frac{K}{\rho_2}\right)[1-e^{-\rho_2(T-h)}]\frac{\sinh\beta h}{\beta}\right\}$$

and

$$B_0=e^{-\rho_2 T} \text{ (a constant independent of } h) \quad (102)$$

where β is given by equation 95. Note that when $K>(\rho_2/2)^2$, β is imaginary. In this case the "cosh" and "sinh" terms are replaced by "cos" and "j sin," respectively. For the continuous system $h=T$, and B_1 becomes:

$$B_1=-2e^{-\frac{\rho_2 T}{2}}\cosh\beta T \quad (103)$$

When β is real, and $K>0$

$$\cosh\beta T<\frac{e^{-\frac{\rho_2 T}{2}}+e^{-\frac{\rho_2 T}{2}}}{2} \quad (104)$$

and

$$B_1>-[1+e^{-\rho_2 T}] \quad (105)$$

Therefore,

$$1+B_1+B_0>0 \quad (106)$$

and the stability conditions become:

$$\left.\begin{aligned}(1-e^{-\rho_2 T})>0\\ 2(1+e^{-\rho_2 T})>0\end{aligned}\right\} \quad (107)$$

These are satisfied for all $\rho_2>0$. Similarly, when β is imaginary B_1 becomes:

$$B_1=-2e^{-\frac{\rho_2 T}{2}}\cos\beta T \quad (108)$$

Substitution of this in equations 89 will show that the stability conditions are again satisfied for all $\rho_2>0$. This means the equivalent continuous system is always stable if $\rho_2>0$, a result which is known from continuous servo theory. Similar methods can be used to derive the stability conditions as $h\to 0$ and the gain K is replaced by K/h. This will be found to agree with the conditions derived directly from the equivalent sampled-data system.

The values of the characteristic coefficients, together with that of the roots z_1 and z_2 are shown in Table II for various values of the pulse width h when

$$\begin{aligned}&K=6\\ &\rho_2=5\\ &T=1 \text{ second}\end{aligned} \quad (109)$$

Fig. 10 shows the continuous response of this system to a unit step input as h is varied. Fig. 11 shows the same except that the gain is increased from six to $K=15.25$. This gives α_1 and α_3 a complex value. Note that for $h=0.4$ second the system has the same rise time as the continuous case but no overshoot. Figs. 12 and 13 show the response of the system defined by equation 109 to a ramp and a sinusoidal input respectively. It is seen that in each case as $h\to T$ the response approaches the equivalent con-

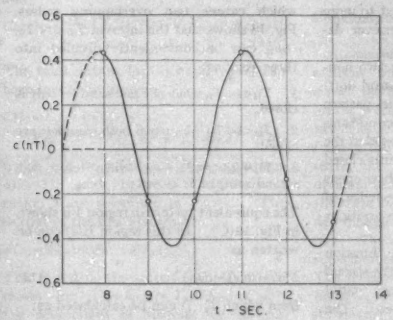


Fig. 15. Plot shows for h = 0.6 the steady-state sinusoidal response at points separated by an integer multiple of the sampling period, lie on a sinusoid having the same frequency as the input sine wave

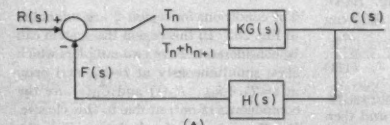


Fig. 16 (left). System with elements in the feedback path

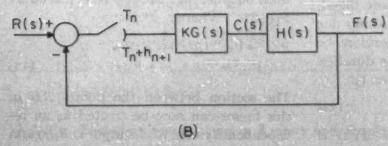
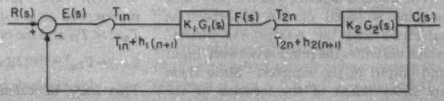


Fig. 17 (below). System with two nonperiodic samplers



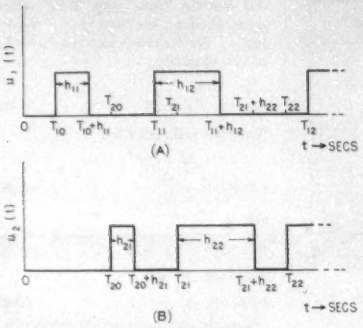


Fig. 18 (left). Pattern of operation for the samplers shown in Fig. 17

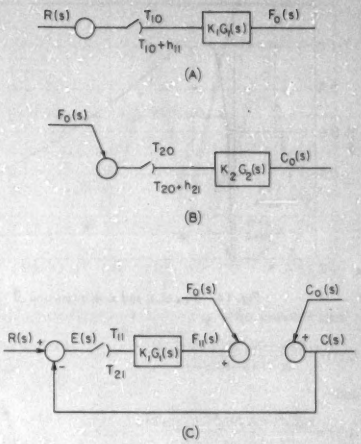


Fig. 19 (right). Equivalent systems of Fig. 17 at different intervals

A — Equivalent system for $0<t<T_{20}$
B — Equivalent system for $T_{20}<t<T_{11}$
C — Equivalent system for $T_{11}<t<T_{21}$

tinous system output as calculated from its Laplace transform. Fig. 14 shows the plot of z_1 and z_2 as h is varied. The product of these is a constant independent of h as required by equations 101 and 102. Fig. 13 shows that the steady-state sinusoidal response is a distorted wave except for $h=T$. However, it can be shown analytically that the steady state output at all points separated by an integer multiple of the sampler period lie on a sinusoid of fixed amplitude and phase which has the same frequency as the input sine wave. This is shown graphically in Fig. 15 for the points $t=nT$, $n=(8, 9, 10, 11, 12, 13)$ when $h=0.6$. The relative phase and amplitude of this wave differs for points not separated by integer multiple of the sampler period. It also varies for different values of pulse width h.

Systems with Elements in Feedback Path

Fig. 16(A) shows a system with elements in the feedback path. This may be redrawn as the unity feedback system shown in Fig. 16(B). Clearly the method described previously can now be used to find $F(s)$. Having obtained this, $C(s)$ can be written as:

$$C(s)=\sum_{n=0}^{\infty} P_{T_n}^{T_n+h_{n+1}}[R(s)-F(s)]KG(s) \qquad (110)$$

$C(s)$ and hence $c(t)$ can be determined since $R(s)$, $F(s)$, and $KG(s)$ are known.

Multiple Sampler Systems

The method described in the preceding sections is based on the superposition principles and the p-transform method which describes the relation between input and output of the sampler. Since these are independent of the location of the sampler it is logical to assume that they can be extended to systems with one or more samplers irrespective of where the samplers are located. To illustrate the procedure a system with two nonperiodic samplers will be considered. The method of analysis can then be extended to more complicated configuration whenever desired.

Fig. 17 shows a system with two nonperiodic samplers. Let $u_1(t)$ and $u_2(t)$ shown in Fig. 18 represent the pattern of operation for the first and second sampler respectively. This indicates that the first sampler closes at T_{10}, remains closed for duration h_{11} seconds, and then opens at $T_{10}-h_{11}$.....etc. The knowledge of these patterns and their relative positions are essential for this analysis.

Consider the output for the duration $0<t<T_{20}$. During this interval the second sampler is open and with zero initial conditions assumed $C(s)$ is zero. The system can therefore be represented as an open-loop system shown in Fig. 19(A). Since $T_{10}+h_{11}$ is less than T_{20}, $F_0(s)$ can be written as:

$$F_0(s)=P_{T_{10}}^{T_{10}+h_{11}}[R(s)]K_1G_1(s) \qquad (111)$$

At $t=T_{20}$ the second sampler closes, remaining closed for h_{21} seconds, and then opens at $T_{20}+h_{21}$. During this period the first sampler is open and therefore the input to the second system is entirely due to $F_0(s)$. The system can therefore be represented as in Fig. 19(B) for duration $T_{20}<t<T_{11}$ and the output written as:

$$C_0(s)=P_{T_{20}}^{T_{20}+h_{21}}[F_0(s)]K_2G_2(s) \qquad (112)$$

This may be calculated since $F_0(s)$ is known. Equations 111 and 112 define the components $F_0(s)$ and $C_0(s)$ for all values of time $t>0$. At $t=T_{11}$ the first sampler closes again for the duration of h_{12} seconds. However, the second sampler also closes at $t=T_{21}$ for h_{22} seconds, which causes two overlapping pulses. Fig. 18 shows that the interval $T_{11}<t<T_{21}+h_{22}$ can be conveniently divided into three regions:

1. $T_{11}<t<T_{21}$ when the first sampler only is closed.
2. $T_{21}<t<T_{11}+h_{12}$ when both samplers are closed.
3. $T_{11}+h_{12}<t<T_{21}+h_{22}$ when only the second sampler is closed.

The equivalent system in region 1 is shown in Fig. 19(C). In this region $E(s)$ can be written as:

$$E(s)=R(s)-C_0(s) \qquad (113)$$

from which $F_{11}(s)$ can be calculated as:

$$F_{11}(s)=P_{T_{11}}^{T_{21}}[R(s)-C_1(s)]K_1G_1(s) \qquad (114)$$

The conditions for region 2 are presented in Fig. 20. In this region the system can be considered to have two samplers which close simultaneously at $t=T_{21}$ and open at $t=T_{11}+h_{12}$. $F_{12}(s)$ and $C_{11}(s)$ are the components of output due to this closure. For the purpose of calculating $C(s)$, the total output, Fig. 20 may be replaced by Fig. 21. Where

$$R_1(s)=P_{T_n}^{\infty}{}_1[R(s)-C_0(s)]-P_{T_n}^{T_{21}+h_{22}}[F_0(s)+F_{11}(s)]K_2G_2(s) \qquad (115)$$

The section between the points AB in this figure can now be treated as an independent system whose input is $R_1(s)$ and

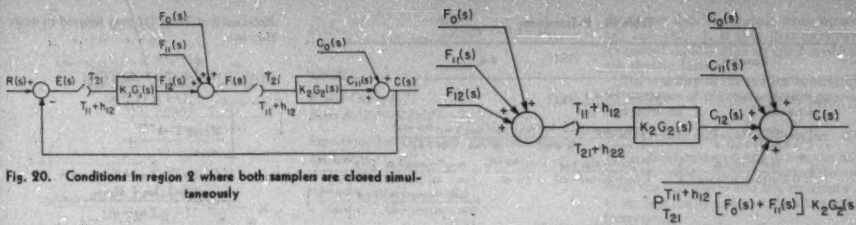


Fig. 20. Conditions in region 2 where both samplers are closed simultaneously

Fig. 22. Equivalent system of Fig. 17 in region 3 where $T_{11}+h_{12}<t<T_{21}$

output $C_{11}(s)$. While both samplers are closed $E(s)$ can be shown to be:

$$E(s)=\frac{R_1(s)}{1+K_1G_1(s)K_2G_2(s)} \qquad (116)$$
$$T_{21}<t<T_{11}+h_{12}$$

Defining a new time axis

$$t_1=(t-T_{21}) \qquad (117)$$

and proceeding as in Figs. 7 and 8 and equations 47 and 48, it can be shown that:

$$F_{12}(s)=e^{-T_{21}s}\times \left\{P_0^{T_{11}+h_{12}-T_{21}}\left[\frac{e^{T_{21}s}R_1(s)}{1+K_1G_1(s)K_2G_2(s)}\right]K_1G_1(s)\right\} \qquad (118)$$

and

$$C_{11}(s)=e^{-T_{21}s}\left[P_0^{T_{11}+h_{12}-T_{21}}[F_{12}(s)]K_2G_2(s)\right] \qquad (119)$$

The total output for $T_{21}<t<T_{11}+h_{12}$ can therefore be written as:

$$C(s)=C_0(s)+C_{11}(s)+P_{T_{21}}^{T_{11}+h_{12}}[F_0(s)+F_{11}(s)]K_2G_2(s) \qquad (120)$$

which can be calculated from equations 112, 118, and 119. The conditions in region 3 are presented in Fig. 22. The output of the first system is

$$F(s)=F_0(s)+F_{11}(s)+F_{12}(s) \qquad (121)$$

This is fed to the second sampler which closes for $T_{11}+h_{12}<t<T_{21}+h_{22}$. $C_{12}(s)$, the component of output due to this closure, can be written as:

$$C_{12}(s)=P_{T_{11}+h_{12}}^{T_{21}+h_{22}}[F_0(s)+F_{11}(s)+F_{12}(s)]K_2G_2(s) \qquad (122)$$

Therefore the total output for $T_{11}+h_{12}<t<T_{21}$ becomes:

$$C(s)=C_0(s)+C_{11}(s)+P_{T_{21}}^{T_{21}+h_{22}}[F_0(s)+F_{11}(s)]K_2G_2(s)+P_{T_{11}+h_{12}}^{T_{21}+h_{22}}[F_{12}(s)]K_2G_2(s) \qquad (123)$$

Proceeding in this way, the output at any instant may be computed. It is clear that this technique can be applied to systems with any number of samplers, operating in any desired sequence. It is reasonable to assume that with periodic samplers certain simplifications will result. The equations for such systems may be derived in a way similar to that outlined for a single periodic sampler.

Conclusions

A general method is presented for the analysis and stability study of finite pulsed feedback systems. This is based on the p-transform and the superposition principle. The p-transform, a particular form of the Laplace transform, is used to derive the Laplace transform of the individual pulses. This takes into account the pulse shape and the pulse width, as is determined by the operation of the sampler. The effect of these pulses are then superimposed to find the total output. The technique is applicable to a general linear system. This is demonstrated by its extension to a system with two non-periodic samplers.

When the samplers are periodic general equations may be developed for the output. It is shown that for an error-pulsed system with a periodic sampler these equations can be expressed in terms of a set of constant coefficients. These may be divided into two categories, 1. the characteristic coefficients, and 2. the input coefficients. The input coefficients are dependent on the system parameter as well as the input function. The number of these is dependent on the order of the input. The characteristic coefficients are independent of the input. They are functions of the pulse width h, the period T, and the pole-zero configurations of $KG(s)$ and $1/1+KG(s)$. For an mth order system there are m^2 such coefficients. The criteria for stability can be established entirely in terms of the relations between these characteristic coefficients. The method can be applied to a system with a general transfer function. The results of this can then be used for the study of a specific system. One such analysis is carried out in the Appendix for a general second-order transfer function. The equations derived can be used to determine the stability of any second-order error pulsed system. The same procedures may be used to derive the equations for a third and higher order systems whenever desired. These can then be tabulated for the purpose of reference.

The pulse width h can be varied from $0\leq h\leq T$. When $h=T$ is substituted in the equations, the output is identical to that which is calculated directly from the Laplace transform. Therefore, a continuous system may be looked upon as a pulsed system whose pulse width is equal to the sampler period. This point of view may be advantageously employed in a new approach to the synthesis of continuous systems. In a similar way the stability conditions and the response of the equivalent sampled-data system is obtained if the pulse width is allowed to

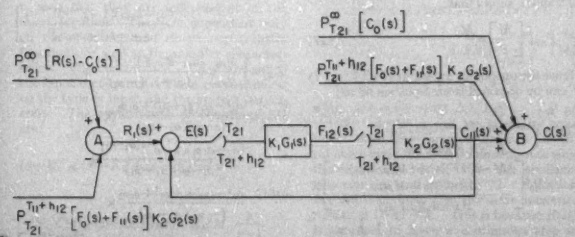


Fig. 21. Equivalent system in region 2 where both samplers are closed simultaneously

Table III. P-Transforms

No.	Function	0-h
1	$P_0^h\left[\frac{a_0}{s+\alpha}\right]$	$\frac{a_0[1-e^{-h(s+\alpha)}]}{(s+\alpha)}$
2	$P_0^h\left[\frac{a_1 s+a_0}{(s+\alpha_1)(s+\alpha_2)}\right]$	$\left[\frac{a_0-a_1\alpha}{\alpha_2-\alpha_1}\frac{1-e^{-h(s+\alpha_1)}}{(s+\alpha_1)}+\frac{a_0-a_1\alpha_2}{\alpha_1-\alpha_2}\frac{1-e^{-h(s+\alpha_2)}}{(s+\alpha_2)}\right]$
3	$P_0^h\left[\frac{a_2 s^2+a_1 s+a_0}{(s+\alpha_1)(s+\alpha_2)(s+\alpha_3)}\right]$	$\left[\frac{a_2\alpha_1^2-a_1\alpha_1+a_0}{(\alpha_2-\alpha_1)(\alpha_3-\alpha_1)}\frac{1-e^{-h(s+\alpha_1)}}{s+\alpha_1}+\frac{a_2\alpha_2^2-a_1\alpha_2+a_0}{(\alpha_1-\alpha_2)(\alpha_3-\alpha_2)}\frac{1-e^{-h(s+\alpha_2)}}{(s+\alpha_2)}+\frac{a_2\alpha_3^2-a_1\alpha_3+a_0}{(\alpha_1-\alpha_3)(\alpha_2-\alpha_3)}\frac{1-e^{-h(s+\alpha_3)}}{(s+\alpha_3)}\right]$
4	$P_0^h\left[\frac{a_1 s+a_0}{s^2}\right]$	$\left[\frac{a_1(1-e^{-hs})-a_0he^{-hs}}{s}+\frac{a_0(1-e^{-hs})}{s^2}\right]$
5	$P_0^h\left[\frac{a_2 s^2+a_1 s+a_0}{s^2(s+\alpha)}\right]$	$\left[\frac{a_2\alpha^2-a_1\alpha+a_0}{\alpha^2}\frac{1-e^{-h(s+\alpha)}}{(s+\alpha)}+\frac{a_1\alpha-a_0}{\alpha^2}\frac{1-e^{-hs}}{s}-\frac{a_0h}{\alpha}\frac{e^{-hs}}{s}+\frac{a_0}{\alpha}\frac{1-e^{-hs}}{s^2}\right]$
6	$P_0^h\left[\frac{a_2 s^2+a_1 s+a_0}{s^2(s+\alpha_1)(s+\alpha_2)}\right]$	$\left\{\frac{a_2\alpha_1^2-a_1\alpha_1+a_0}{\alpha_1^2(\alpha_2-\alpha_1)}\frac{1-e^{-h(s+\alpha_1)}}{(s+\alpha_1)}+\frac{a_2\alpha_2^2-a_1\alpha_2+a_0}{\alpha_2^2(\alpha_1-\alpha_2)}\frac{1-e^{-h(s+\alpha_2)}}{(s+\alpha_2)}+\frac{a_1\alpha_1\alpha_2-a_0(\alpha_1+\alpha_2)}{(\alpha_1\alpha_2)^2}\frac{1-e^{-hs}}{s}+\frac{a_0}{\alpha_1\alpha_2}\frac{1-e^{-hs}}{s^2}-\frac{a_0h}{\alpha_1\alpha_2}\frac{e^{-hs}}{s}\right\}$
7	$P_0^h\left[\frac{a_1 s+a_0}{(s^2+\omega^2)}\right]$	$\left[\frac{a_0-ja_1\omega}{-2j\omega}\frac{1-e^{-h(s+j\omega)}}{(2+j\omega)}+\frac{a_0+ja_1\omega}{2j\omega}\frac{1-e^{-h(s-j\omega)}}{(s-j\omega)}\right]$
8	$P_0^h\left[\frac{a_2 s^2+a_1 s+a_0}{(s^2+\omega^2)(s+\alpha)}\right]$	$\left\{\frac{a_2\alpha^2-a_1\alpha+a_0}{\alpha^2+\omega^2}\frac{1-e^{-h(s+\alpha)}}{(s+\alpha)}+\frac{(a_0-a_2\omega^2)-ja_1\omega}{-2j\omega(\alpha-j\omega)}\frac{1-e^{-h(s+j\omega)}}{(s+j\omega)}+\frac{(a_0-a_2\omega^2)+ja_1\omega}{2j\omega(\alpha+j\omega)}\frac{1-e^{-h(s-j\omega)}}{(s-j\omega)}\right\}$
9	$P_0^h\left[\frac{a_2 s^2+a_1 s+a_0}{(s^2+\omega^2)(s+\alpha_1)(s+\alpha_2)}\right]$	$\left\{\frac{a_2\alpha_1^2-a_1\alpha_1+a_0}{(\alpha_1^2+\omega^2)(\alpha_2-\alpha_1)}\frac{1-e^{-h(s+\alpha_1)}}{(s+\alpha_1)}+\frac{a_2\alpha_2^2-a_1\alpha_2+a_0}{(\alpha_2^2+\omega^2)(\alpha_1-\alpha_2)}\frac{1-e^{-h(s+\alpha_2)}}{(s+\alpha_2)}+\frac{(a_0-a_2\omega^2)-ja_1\omega}{-2j\omega[(\alpha_1\alpha_2-\omega^2)-j\omega(\alpha_1+\alpha_2)]}\frac{1-e^{-h(s+j\omega)}}{(s+j\omega)}+\frac{(a_0-a_2\omega^2)+ja_1\omega}{2j\omega[(\alpha_1\alpha_2-\omega^2)+j\omega(\alpha_1+\alpha_2)]}\frac{1-e^{-h(s+j\omega)}}{(s-j\omega)}\right\}$

approach zero and the gain is increased inversely as h so that the product of h times the new gain is always equal to the fixed open loop gain.

Assumptions 1 and 2 for the feedback systems considered can be relaxed when necessary. The initial conditions can always be represented by an additional component $C_f(s)$. This component can then be treated exactly the same way as $C_0(s)$.....etc. Assumption 2 can be modified to include systems whose numerator and denominator polynomials are of the same order. In this case the same method can be used except that a distinction must be made between $c(nT)^-$, $c(nT)^+$, $c(nT+h)^-$, and $c(nT+h)^+$.

In concluding it is well to mention that the presence of the sampler is not an essential requirement for the application of this technique. Any linear system with finite pulse inputs at its various stages lends itself to analysis by this method irrespective of how these pulses are generated.

Appendix

A general second order transfer function may be represented by

$$KG(s)=\frac{a_1 s+a_0}{(s+\rho_1)(s+\rho_2)} \tag{124}$$

where it is assumed $\rho_1 \neq \rho_2$. If $\rho_1=\rho_2$ then the same procedure can be used except that the values of the residues are different. Consider the error-pulsed system shown in Fig. 9. This has a periodic sampler with period T and pulse width h. The stability of such a system is independent of the output; however, for the purpose of illustration a step input of magnitude M will be assumed. Therefore

$$R(s)=\frac{M}{s} \tag{125}$$

and

$$\frac{1}{1+KG(s)}=\frac{(s+\rho_1)(s+\rho_2)}{s^2+(\rho_1+\rho_2+a_1)s+(\rho_1\rho_2+a_0)}=\frac{(s+\rho_1)(s+\rho_2)}{(s+\alpha_1)(s+\alpha_2)} \tag{126}$$

where α_1 and α_2 are the poles of $1/1+KG(s)$ and can have complex values. Relation 1 of Table I shows that:

$$e^{nTs}P_{nT}^{\infty}\left[\frac{M}{s}\right]=\frac{M}{s} \tag{127}$$

Hence the output for duration $nT<t\leq(n+1)T$ can be obtained from equation 68 as:

$$C(s)=e^{-nTs}\left\{\frac{c(nT)s+A_0(nT)}{(s+\rho_1)(s+\rho_2)}+P_0^h\left[\frac{[M-c(nT)]s^2+[M(\rho_1+\rho_2)-A_0(nT)]s+M\rho_1\rho_2}{s(s+\alpha_1)(s+\alpha_2)}\right]\times\frac{a_1 s+a_0}{(s+\rho_1)(s+\rho_2)}\right\} \tag{128}$$

Relation 3 of Table III may be used to write this as:

$$C(s)=e^{-nTs}\left\{\frac{c(nT)s+A_0(nT)}{(s+\rho_1)(s+\rho_2)}+\left[\frac{M\rho_1\rho_2}{\alpha_1\alpha_2}\frac{1-e^{-hs}}{s}+\frac{[M-c(nT)]\alpha_1^2-[M(\rho_1+\rho_2)-A_0(nT)]\alpha_1+M\rho_1\rho_2}{(-\alpha_1)(\alpha_2-\alpha_1)}\times\frac{1-e^{-h(s+\alpha_1)}}{s+\alpha_1}+\frac{[M-c(nT)]\alpha_2^2-[M(\rho_1+\rho_2)-A_0(nT)]\alpha_2+M\rho_1\rho_2}{(-\alpha_2)(\alpha_1-\alpha_2)}\times\frac{1-e^{-h(s+\alpha_2)}}{s+\alpha_2}\right]\frac{a_1 s+a_0}{(s+\rho_1)(s+\rho_2)}\right\} \tag{129}$$

The required difference equations can now be obtained from the following identity:

$$P_{(n+1)T}^{\infty}[\text{equation 129}]=e^{-(n+1)Ts}\times\left[\frac{c(n+1)Ts+A_0(n+1)T}{(s+\rho_1)(s+\rho_2)}\right] \tag{130}$$

Utilizing relation 6 of Table I, the left hand side of this equation may be written as:

$$P_{(n+1)T}^{\infty}[\text{equation 129}]=e^{-(n+1)Ts}\times\left\{[L_c{}^1c(nT)+L_0{}^1A_0nT)+L_R{}^1M]\times\frac{e^{-\rho_1 T}}{s+\rho_1}+[L_c{}^2c(nT)+L_0{}^2A_0(nT)+L_R{}^2M]\frac{e^{-\rho_2 T}}{s+\rho_2}\right\} \tag{131}$$

where

$$L_c{}^1=\left[\frac{-\rho_1}{\rho_2-\rho_1}+\frac{\alpha_1}{\alpha_2-\alpha_1}\times\frac{(a_0-a_1\rho_1)(1-e^{-h(\alpha_1-\rho_1)})}{(\rho_2-\rho_1)(\alpha_1-\rho_1)}+\frac{\alpha_2}{\alpha_1-\alpha_2}\times\frac{(a_0-a_1\rho_1)(1-e^{-h(\alpha_2+\rho_1)})}{(\rho_2-\rho_1)(\alpha_2-\rho_1)}\right]$$

$$L_0{}^1=\left[\frac{1}{\rho_2-\rho_1}-\frac{1}{(\alpha_2-\alpha_1)}\times\frac{(a_0-a_1\rho_1)(1-e^{-h(\alpha_1-\rho_1)})}{(\rho_2-\rho_1)(\alpha_1-\rho_1)}-\frac{1}{\alpha_1-\alpha_2}\times\frac{(a_0-a_1\rho_1)(1-e^{-h(\alpha_2-\rho_1)})}{(\rho_2-\rho_1)(\alpha_2-\rho_1)}\right]$$

$$L_R{}^1=\left[\frac{\rho_1\rho_2}{\alpha_1\alpha_2}\frac{(a_0-a_1\rho_1)(1-e^{-h\rho_1})}{(\rho_2-\rho_1)(-\rho_1)}+\frac{\alpha_1{}^2-\alpha_1(\rho_1+\rho_2)+\rho_1\rho_2}{(-\alpha_1)(\alpha_2-\alpha_1)}\times\frac{(a_0-a_1\rho_1)(1-e^{-h(\alpha_1-\rho_1)})}{(\rho_2-\rho_1)(\alpha_1-\rho_1)}+\frac{\alpha_2{}^2-\alpha_2(\rho_1+\rho_2)+\rho_1\rho_2}{(-\alpha_2)(\alpha_1-\alpha_2)}\times\frac{(a_0-a_1\rho_2)(1-e^{-h(\alpha_2-\rho_1)})}{(\rho_2-\rho_1)(\alpha_2-\rho_1)}\right]$$

$$L_c{}^0=\left[-\frac{\rho_2}{\rho_1-\rho_2}+\frac{\alpha_1}{\alpha_2-\alpha_1}\times\frac{(a_0-a_1\rho_2)(1-e^{-h(\alpha_1-\rho_2)})}{(\rho_1-\rho_2)(\alpha_1-\rho_2)}+\frac{\alpha_2}{\alpha_1-\alpha_2}\times\frac{(a_0-a_1\rho_2)(1-e^{-h(\alpha_2-\rho_2)})}{(\rho_1-\rho_2)(\alpha_2-\rho_2)}\right]$$

$$L_0{}^0=\left[\frac{1}{\rho_1-\rho_2}-\frac{1}{\alpha_2-\alpha_1}\times\frac{(a_0-a_1\rho_2)(1-e^{-h(\alpha_1-\rho_2)})}{(\rho_1-\rho_2)(\alpha_1-\rho_2)}-\frac{1}{\alpha_1-\alpha_2}\times\frac{(a_0-a_1\rho_2)(1-e^{-h(\alpha_2-\rho_2)})}{(\rho_1-\rho_2)(\alpha_2-\rho_2)}\right]$$

$$L_R{}^0=\left[\frac{\rho_1\rho_2}{\alpha_1\alpha_2}\frac{(a_0-a_1\rho_2)(1-e^{h\rho_2})}{(\rho_1-\rho_2)(-\rho_2)}+\frac{\alpha_1{}^2-\alpha_1(\rho_1+\rho_2)+\rho_1\rho_2}{(-\alpha_1)(\alpha_2-\alpha_1)}\times\frac{(a_0-a_1\rho_2)(1-e^{-h(\alpha_1-\rho_2)})}{(\rho_1-\rho_2)(\alpha_1-\rho_2)}+\frac{\alpha_2{}^2-\alpha_2(\rho_1+\rho_2)+\rho_1\rho_2}{(-\alpha_2)(\alpha_1-\alpha_2)}\times\frac{(a_0-a_1\rho_2)(1-e^{-h(\alpha_2-\rho_2)})}{(\rho_1-\rho_2)(\alpha_2-\rho_2)}\right] \quad (132)$$

Equation 131 may be placed on a common denominator and written as:

$$P_{(n+1)T}{}^{\infty}[\text{equation } 129]=e^{-(n+1)Ts}\times\frac{\left\{\begin{array}{l}[K_c{}^0c(nT)+K_c{}^0A_0(nT)+K_R{}^0M]s+\\ [K_0{}^0c(nT)+K_0{}^0A_0(nT)+K_R{}^0M]\end{array}\right\}}{(s+\rho_1)(s+\rho_2)} \quad (133)$$

where

$$\begin{aligned}K_c{}^c&=L_c{}^1e^{-\rho_1T}+L_c{}^2e^{-\rho_2T}\\ K_0{}^c&=L_0{}^1e^{-\rho_1T}+L_0{}^2e^{-\rho_2T}\\ K_R{}^c&=L_R{}^1e^{-\rho_1T}+L_R{}^2e^{-\rho_2T}\\ K_c{}^0&=\rho_1L_c{}^1e^{-\rho_1T}+\rho_1L_c{}^2e^{-\rho_2T}\\ K_0{}^0&=\rho_2L_0{}^1e^{-\rho_1T}+\rho_1L_0{}^2e^{-\rho_2T}\\ K_R{}^0&=\rho_2L_R{}^1e^{-\rho_1T}+\rho_1L_R{}^2e^{-\rho_2T}\end{aligned} \quad (134)$$

Equating equations 130 and 133, the following difference equations are obtained:

$$c(n+1)T=K_c{}^cc(nT)+K_0{}^cA_0(nT)+K_R{}^cM \quad (135)$$

$$A_0(n+1)T=K_c{}^0c(nT)+K_0{}^0A_0(nT)+K_R{}^0M$$

The constant coefficients $K_c{}^c$, $K_0{}^c$, $K_c{}^0$, $K_0{}^0$ are called the characteristic coefficients. It is seen that they are independent of the input function. They are dependent only on the open-loop and closed-loop transfer functions, as well as the sampler characteristics. The constants $K_R{}^c$, $K_R{}^0$ are called the input coefficients. These are dependent on the type of input and the system parameters. The z-transforms of equations 135 are:

$$(z-K_c{}^c)C(z)-K_0{}^cA_0(z)=K_R{}^cM\frac{z}{z-1} \quad (136)$$

$$-K_c{}^0C(z)+(z-K_0{}^0)A_0(z)=K_R{}^0M\frac{z}{z-1}$$

The characteristic equation is given by:

$$\Delta(z)=\begin{vmatrix}z-K_c{}^c & -K_0{}^c\\ -K_c{}^0 & z-K_0{}^0\end{vmatrix}=(z^2+B_1z+B_0) \quad (137)$$

where

$$\begin{aligned}B_1&=-(K_c{}^c+K_0{}^0)\\ B_0&=(K_c{}^cK_0{}^0-K_0{}^cK_c{}^0)\end{aligned} \quad (138)$$

Equations for $C(z)$ and $A_0(z)$ can also be written as:

$$C(z)=\frac{\begin{vmatrix}K_R{}^cM\frac{z}{z-1} & -K_0{}^c\\ K_R{}^0M\frac{z}{z-1} & z-K_0{}^0\end{vmatrix}}{\Delta(z)}=\frac{Mz[K_R{}^cz-(K_0{}^0K_R{}^c-K_0{}^cK_R{}^0)]}{(z-1)(z^2+B_1z+B_0)} \quad (139)$$

and

$$A_0(z)=\frac{\begin{vmatrix}z-K_c{}^c & K_R{}^cM\frac{z}{z-1}\\ -K_c{}^0 & K_R{}^0M\frac{z}{z-1}\end{vmatrix}}{\Delta(z)}=\frac{Mz[K_R{}^0z-(K_c{}^cK_R{}^0-K_c{}^0K_R{}^0)]}{(z-1)(z^2+B_1z+B_0)} \quad (140)$$

The inverse of equations 139 and 140 will give the values of $c(nT)$ and $A_0(nT)$. Note that the M appears as a multiplier indicating a direct proportionality between the input and output amplitudes. The continuous response can be evaluated in the following way. $C_h(s)$, the component of output valid for $nT<t\leq(nT+h)$ can be written in accordance with equation 70 as:

$$C_h(s)=e^{-nTs}\left\{\frac{c(nT)s+A_0(nT)}{(s+\alpha_1)(s+\alpha_2)}+\frac{M(a_1s+a_0)}{s(s+\alpha_1)(s+\alpha_2)}\right\} \quad (141)$$
$$nT<t\leq(nT+h)$$

The inverse Laplace of this equation describes the continuous output for the interval $nT<t\leq nT+h$, i.e., while the sampler is closed. This can be calculated since $c(nT)$ and $A_0(nT)$ are known. The second component of output valid for $nT+h<t\leq(n+1)T$ can be derived from equation 131 and is:

$$C_T(s)=e^{-(nT+h)s}\times\left\{[L_c{}^1c(nT)+L_0{}^1A_0(nT)+L_R{}^1M]\times\frac{e^{-\rho_1h}}{s+\rho_1}+[L_c{}^2c(nT)+L_0{}^2A_0(nT)+L_R{}^2M]\frac{e^{-\rho_2h}}{s+\rho_2}\right\} \quad (142)$$
$$nT+h<t\leq(n+1)T$$

where the coefficients $L_c{}^1$, $L_0{}^1$ ..... $L_R{}^2$ are already calculated and are given by equations 132. The inverse Laplace of this equation gives the continuous response while the sampler is open. Note the similarity between this and equation 72. Equations 141 and 142 must give the same numerical results at $t=nT+h$. This is because $KG(s)$ is assumed to have a continuous step response. The simplification which results from division of $C(s)$ into its two components $C_h(s)$ and $C_T(s)$ is evident from equations 141 and 142. It is seen that $c_T(t)$ can be computed rapidly since all the L coefficients are already known.

The stability of the system is determined by the roots of the characteristic equations. Let these be

$$\Delta(z)=z^2+B_1z+B=(z-z_1)(z-z_2) \quad (143)$$

For a stable system z_1 and z_2 must lie within the unit circle. It is convenient to transform the unit circle into the imaginary axis of the w-plane by means of the bilinear transformation,[11]

$$z=\frac{w+1}{w-1} \quad (144)$$

This transforms the inside of the unit circle into the left half of w-plane. Therefore, a stable system must have all its roots in this half of the w-plane. Routh's criteria can now be applied to the transformed $\Delta(z)$ to obtain the stability conditions in terms of the B coefficients. In this particular case $\Delta(z)$ transforms to:

$$\left(\frac{w+1}{w-1}\right)^2+B_1\left(\frac{w+1}{w-1}\right)+B_0=(1+B_1+B_0)w^2+2(1-B_0)w+(1-B_1+B_0) \quad (145)$$

and the stability conditions become:

$$\begin{aligned}\frac{1-B_0}{1+B_1+B_0}&>0\\ \frac{1-B_1+B_0}{1+B_1+B_0}&>0\end{aligned} \quad (146)$$

Thus, given a specific second order system the B coefficients can be calculated from equations 132, 134, and 138. These may then be substituted in conditions 146 to determine the stability of the system prior to calculating the system response.

The same method can be used to derive the equations for a third and higher order systems. Having once derived these in terms of a general transfer function, they can be used repeatedly to evaluate the stability and response of specific systems.

References

1. The Analysis of Sampled-Data Systems J. R. Ragazzini, L. A. Zadeh. *AIEE Transactions*, vol. 71, pt. II, 1952, pp. 225–34.

2. Analysis and Synthesis of Sampled-Data Control Systems, Eliahu I. Jury. *Ibid.*, vol. 73, pt. I, 1954, pp. 332–46.

3. The Pulse Transfer Function and Its Application to Sampling Servo-Systems, R. H. Barker. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. IV, 1952, pp. 302–17.

4. Synthesis and Critical Study of Sampled-Data Control Systems, Eliahu I. Jury. *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 141–51.

5. Analysis of Linear Sampled-Data Systems with Finite Pulse Width (Open Loop), G. Farmanfarma. *Electronics Research Laboratory Report*, Series no. 60, Issue no. 157, University of California, Berkeley, Calif., Apr. 1956.

6. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, G. Farmanfarma. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 808–19.

7. Application of the β-Transform Method to Analysis of Multiple Samples Systems (Open Loop), G. Farmanfarma. *Electronics Research Laboratory Report*, Series no. 60, Issue no. 174, University of California, Berkeley, Calif., Jan. 1957.

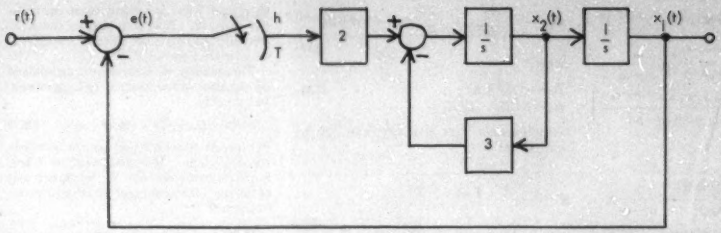


Fig. 23. System diagram

8. Analysis of Multiple Sampler Systems with Finite Pulse Width, Open Loop, **G. Farmanfarma.** *AIEE Transactions*, vol. 77, pt. II, Mar. 1958, pp. 20–28.

9. Experimental Study of a Sampled Feedback System with Finite Pulse Width, **I. W. Metzger.** *M.S. Thesis*, University of California, Berkeley, Calif., May 1957.

10. Transients in Linear Systems (book), vol. I, **M. F. Gardner, J. L. Barnes.** John Wiley & Sons, Inc., New York, N. Y., 1953, pp. 275–78.

11. Dynamics of Automatic Controls (book), **R. C. Oldenbourg, H. Sartorius** (English translation by H. L. Mason). American Society of Mechanical Engineers, New York, N. Y., 1948, chap. V, pp. 190–93.

Discussion

R. E. Kalman (International Business Machines Corporation, Yorktown, N. Y.) and **J. E. Bertram** (Columbia University, New York, N. Y.): The author is to be commended for obtaining an exact solution to this problem which requires no more effort in the computational sense than some of the approximate solutions which have been proposed.

The discussers feel, however, that introduction of the concept of the "state" of a dynamic system permits a more orderly approach to the problem and results in considerable savings in labor in most cases.[1] By the state of a dynamic system is meant here a set of numbers (called state variables) which contain all the information of the past history of the system which is required for the calculation of the future behavior of the system for a known input.

To illustrate this method, consider Fig. 23. This is the same system as that shown in Fig. 9 if we let

$$KG(s)=\frac{2}{s(s+3)}$$

The sampler shown in Fig. 23 and in Fig. 9 operates every T seconds in such a way that for h seconds the output of the error detector is applied directly to $KG(s)$, after which the switch is opened for $T-h$ seconds so that the input to $KG(s)$ is zero. Thus during the first time-interval $kT\leq t<kT+h$ we have a closed-loop system while during the second time interval $kT+h\leq t<(k+1)T$ $KG(s)$ operates open loop with its input zero.

During the first time-interval, we see from Fig. 23 that the system is governed by the differential equations

$$\begin{aligned} dx_1/dt &= x_2 \\ dx_2/dt &= -2x_1-3x_2+2r(t) \end{aligned} \qquad (147)$$

During the second time-interval, the system equations are

$$\begin{aligned} dx_1/dt &= x_2 \\ dx_2/dt &= -3x_2 \end{aligned} \qquad (148)$$

The stability of the system, because of linearity, does not depend on $r(t)$. Hence let $r(t)=0$. Then, using standard vector-matrix notation, equations 147 and 148 may be written in the form

$$d\mathbf{x}/dt=A\mathbf{x} \qquad (149)$$

The vector $\mathbf{x}=(x_1, x_2)$ is evidently the state of the system. During the first time-interval, the 2×2 matrix A is given by

$$A_1=\begin{Vmatrix} 0 & 1 \\ -2 & -3 \end{Vmatrix}$$

During the second time-interval, A is given by

$$A_2=\begin{Vmatrix} 0 & 1 \\ 0 & -3 \end{Vmatrix}$$

As is well known, the unique solution of equation 149 for any initial state $\mathbf{x}(t_0)$ at $t=t_0$ is given by the vector function

$$\mathbf{x}(t)=(\exp A(t-t_0))\mathbf{x}(t_0) \qquad (150)$$

One particular definition of the exponential function of a matrix is

$$\exp At=\sum_{k=0}^{\infty}A^kt^k/k!\ (A^0=\text{unit matrix}) \qquad (151)$$

Using these results, it follows that the states of the system in Fig. 23 are governed by the transition equation

$$\mathbf{x}((k+1)T)=(\exp A_2(T-h))(\exp A_1h)\mathbf{x}(kT) \qquad (152)$$

Equation 152 is a system of two simultaneous difference equations which shows how the state of the system at time $t=(k+1)T$ can be computed from the knowledge of the state of the system at time $t=kT$, provided the input is zero.

The sampled-data system in Fig. 23 is stable if and only if all roots of the characteristic equation

$$\det[(\exp A_2(T-h))(\exp A_1h)-zI]=0 \qquad (153)$$

are less than unity in absolute value. In equation 153, det means the determinant of a matrix and I is the unit matrix.

The matrices in 152 and 153 can be computed numerically to any desired degree of accuracy from the rapidly converging infinite series 151. It is also a simple matter to express the matrix $\exp At$ in closed form as an explicit function of t.[1]

Similarly to the foregoing derivation, it is possible to include the effect of the input $r(t)$ in the analysis and also to obtain simple expressions for the behavior of the system between sampling instants.

Reference

1. General Synthesis Procedure for Computer Control of Single and Multiloop Linear Systems, **R. E. Kalman, J. E. Bertram.** *AIEE Special Publication T-101*, "Conference on Computers in Control Systems," May 1958, pp. 130–37.

G. Farmanfarma: The author expresses his sincere thanks to Messrs. Kalman and Bertram for their interest and the discussion they have presented.

The concept of the "state" of a dynamic system as defined by the discussers is of course well known in principle. The state of a linear system is uniquely defined, at any instant of time, by a set of numbers from which the function and its derivatives can be obtained. Such a set of numbers are denoted by $c(nT)$, $A_{m-1}(nT)$, ... $A_1(nT)$, and $A_0(nT)$ in the paper and are referred to in the paragraph following equation 67.

One of the main objectives of this paper is to derive a set of simultaneous first order difference equations which relate the nth and the $(n+1)$th term. These are shown, for the general periodic case, in equations 79 and are derived by carrying out specified operations on the Laplace transform of the function. The boundary conditions are then equated directly in the time domain as indicated by equations 74 through 79 to obtain the required set of equations.

In the discussion presented, the same equations are derived from the differential equations in the time domain. Thus, for the second-order system considered, equation 152 of the discussion may be readily compared with equation 135 of the paper. This solution is of course equally valid and perhaps simpler in certain cases. However, it is the author's belief that in the general case the Laplace transform approach will be found more convenient because it is then possible to take advantage of general tables. For example, the task of solving any second order system now is purely algebraic in view of the general equations developed in the appendix.

The suggestion of approximating the coefficients by a converging infinite series is a welcome one. However, it is essential to know the error which is introduced in the final results by such an approximation. This requires further investigation, and it is hoped that Messrs. Kalman and Bertram will be able to guide us on this matter in the future.

The Financial Return on Railway Electrification—An Approximate Method for Its Determination

H. F. BROWN
FELLOW AIEE

H. E. MARMAROS
MEMBER AIEE

IN THE PAST, railway electrification was applied in the United States to overcome limitations or difficulties generally involved with steam locomotive operation. Today, electrification when it is to be applied, must justify itself economically in comparison with diesel electric operation which has already largely superseded steam operation.

The future type of electrification envisioned is no longer burdened with investment in power supply, transmission lines or conversion equipment. It is powered directly from existing power supply systems and transmission networks with commercial frequency directly on the contact system. This system has the endorsement of the electrical section of the Association of American Railroads.

Under what circumstances will electric operation of a railroad be cheaper than diesel operation? A simple chart (Fig. 1) has been prepared by means of which any railroad can determine whether electrification offers any financial advantage for its operation, and if so how much. To use this chart a few basic factors must be known:

1. How many locomotives and of what horsepower capacity are now required to handle the traffic over the mileage being considered for electrification?

2. What is the total track mileage that would be involved?

3. What is the average annual mileage made per locomotive?

From the total locomotive horsepower and the track mileage involved, the rail horsepower (rhp) per mile can be calculated. The "rail horsepower" is generally taken as 82% of the rated diesel engine horsepower.

Two examples will serve to illustrate the use of the chart:

1. Railroad *X* has a single track main line 400 miles long. Approximately 500 miles of track would require a contact system. This line is now operated by 40-1,750 hp (engine rating) road diesels, which make on an average 100,000 miles per year each. Should electrification be considered for this railroad?

The rhp per mile is first calculated. This is $\frac{40 \times 1{,}750 \times 82\%}{500}$ or 115 rhp per mile. On the chart, Fig. 1, locate the point P_1 at the intersection of 115 rhp per mile and the mileage, 100,000 per year. This point is well below any of the curves on the chart which show "per-cent return" on the additional investment required. Thus, electrification does not appear to be economically feasible for this railroad.

2. Railroad *Y* has a double track main line 900 miles long. Approximately 2,000 miles of track would require a contact system. The present road motive power assigned to this route consists of 625 diesels averaging 1,500 hp each. The average mileage made annually by each unit is 75,000.

This "rail horsepower per mile" is: $\frac{625 \times 1{,}500 \times 82\%}{2{,}000}$ or 385 rhp per mile

On the chart, this value of rhp per mile intersects the 75,000 annual mileage line at P_2 which is very nearly 24% return on the additional investment required for electrification.

A detailed study of all the advantages of electrification seems to be indicated in this case provided the present motive power is approaching the age for renewal, and provided electric power is available where needed along this railroad at a cost competitive with the cost of diesel fuel.

The chart has been prepared from the fundamental concept that the return *R* on any increase in investment made for electric operation over diesel operation is, in per cent,

$$R = \frac{\text{difference in operating costs and fixed charges}}{\text{difference in total investment}} \times 100$$

All of the factors making up these differences may be expressed as functions of one or more of the following basic factors:

A. The miles of track to be considered as electrified to handle the required traffic. This determines the additional investment required for the contact line and any other costs related thereto, as well as the magnitude of the related maintenance and operating costs, and fixed charges.

B. The number and capacity (or rating) of the present motive power units, giving the total horsepower required to move the traffic with either form of motive power. A convenient unit for comparing the performance of various types of motive power is "1,000 continuous horsepower measured at the rim of the driving wheels," referred to in this discussion as "1,000 rhp." It will be assumed in this discussion that the same number of 1,000 rhp units will be required to perform the same service *at the same speed* with either form of motive power. The difference in investment in motive power and the fixed charges thereon are determined by *B*.

C. The annual mileage made by the locomotive units as defined in *B*, in performing their work. Motive power operating costs are determined by both *B* and *C*.

It appears possible therefore by making certain assumptions as to the values of *A*, *B*, and *C*, to arrange these in a formula such that the return *R* on the difference in investment between the two types of operation, electric versus diesel, is expressed by some combination of functions of *A*, *B*, and *C*; or

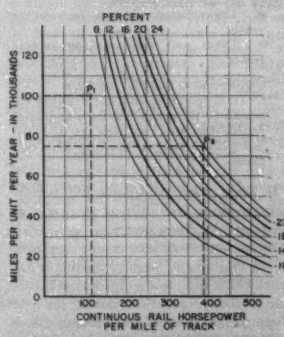


Fig. 1. Per-cent return on additional investment for electrification in terms of traffic density

Paper 58-32, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 16, 1957; made available for printing November 4, 1957.

H. F. BROWN and H. E. MARMAROS are with Gibbs & Hill, Inc., New York, N. Y.

$R=f(A, B, C)$

A little thought will show that these three factors also determine "Traffic Density."

The Difference in Investment

For electric operation, the investment involved consists of:

1. The cost of the contact system (a function of A).
2. The cost of the power control and switching stations (a function of A).
3. The cost of signal and communication changes (a function of A).
4. The cost of the motive power (a function of B).

For diesel operation, the investment consists of the cost of motive power only (a function of B).

Let K_1 equal the cost per mile of all of the investment in 1, 2 and 3 of the preceding for electric operation and let K_2 be the difference in cost between diesel and electric motive power per 1,000-rhp unit. Then the total difference in investment becomes K_1A-K_2B. K_2 is negative because electric motive power should cost less than diesel on a rail horsepower basis.

Estimates are available for the values of K_1 and K_2. The annual return R is to be calculated on this difference in investment.

The Difference in Operating Costs

The major items of operating expense wherein substantial differences may be found as between the two methods of operation can be divided into two groups:

1. *Favorable to Electric Operation*
   a. Locomotive maintenance expense
   b. Engine house expense.
   c. Lubrication expense.
   d. Depreciation charges on motive power.
   e. Taxes and insurance on motive power.
2. *Unfavorable to Electric Operation*
   f. The operating and maintenance expenses of the contact system and associated facilities, as well as the depreciation charges thereon and any taxes.

In some states additions and betterments made to fixed railroad property are not subject to direct taxation.

In the not-too-distant future there will be one more major item of operating expense which will show a differential favorable to electric operation, namely the cost of electric power versus the cost of diesel fuel oil. At the present time in the United States these costs are so very nearly equal for the usual load factors that this factor can be disregarded in the present discussion. The trends of both of these energy costs, however, are too well known to need amplification.

In combining and comparing all these costs, let K_D be the sum of the operating costs a, b, c enumerated previously for diesel operation reduced to a "per 1,000 rhp-mile" unit cost, and let K_E be the similar unit costs for electric operation. Also let K_3 be the difference in these unit costs, which is always in favor of electric operation. $K_3=K_D-K_E$.

The magnitude of the savings in these costs with electric operation is therefore the unit savings K_3 multiplied by the product of the number of "units" of motive power B and the annual mileage C, or K_3BC.

It will cost a certain amount, annually, to maintain and operate the electric contact system and power supply control. Let K_4 be this cost per mile per year. The total annual cost of this item becomes K_4A which must be subtracted from the operating savings K_3BC already indicated. K_4 is therefore negative.

Statistical data are available for the determination of the value of K_3 and K_4.

Fixed Charges Against Operation

The depreciation charges for motive power is under Interstate Commerce Commission accounting procedure an annual charge against operating expense. Since electric motive power is known to have a longer economic life than diesel power, and since the investment is less for the same total horsepower, the depreciation charges will be substantially less with electric motive power.

Taxes on both types of motive power are generally based on the average valuation over the life. These will also be less with electric operation, since the rate would be the same with either type of motive power, with a smaller investment for the electric. Any insurance factor could also be included in this differential.

If K_5 is the total differential in these costs for motive power, this factor becomes K_5B. From this must be subtracted the similar charge for depreciation (and taxes, if any) against the investment in the line facilities.

Let K_6 be the sum of depreciation, taxes and insurance charges against the line investment per mile. This factor then becomes K_6A. The determination of K_5 and K_6 can be made from known values and rates.

For the same service, i.e., the same ton miles moved per year at the same speed, the total savings with electric operation becomes

$$K_3BC-K_4A+K_5B-K_6A$$

or

$$K_3BC+K_5B-(K_4+K_6)A$$

The return R in per cent on the investment then becomes

$$R=\frac{K_3BC+K_5B-(K_4+K_6)A}{K_1A-K_2B}\times 100$$

Because data are available for determining the values of K_1, K_2, K_3, K_4, K_5 and K_6, it now becomes possible by using C and B/A as co-ordinates to show R graphically for various values of A, B, and C as in Fig. 2.

The Fig. 1 chart has been made from the various values of the constants K_1 to K_6, inclusive, determined from carefully prepared and conservative estimates as to investment costs, and from data relative to known maintenance and other operating costs taken from a number of electrically operated and diesel operated railways in this country.

It is not intended that this chart should be regarded as the final answer to the economic problems involved with railway electrification. It simply serves as a general guide to show what traffic density presents economic possibilities

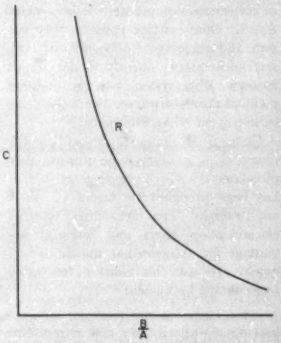


Fig. 2. Return on investment as a function of average annual mileage and rail horsepower per mile

A=Track mileage under consideration for electrification

B=Total rail horsepower now being used to operate over the mileage A

C=Average annual mileage made by the present motive power in this service

$\frac{B}{A}$=Rail horsepower per mile

R=Return on the added investment for electrification after depreciation, state and local taxes, insurance, and operating expenses but before other financial charges

as well as where the traffic alone will not warrant the additional investment.

The financial return can be greater, generally, than that indicated because of operation factors more favorable with electrification which are not direct functions of the traffic density.

One of these factors is increased speed. In many cases increased speeds would increase the financial return because:

1. Higher speeds will attract more high revenue traffic now lost to competitive transportation agencies.

2. Higher speeds would reduce the investment in car equipment, and motive power although this might require more electric horsepower than diesel in some cases.

3. Higher speeds would certainly cut down per diem expense for cars.

In some cases the economic effects of speed can be evaluated for a particular situation with much the same reasoning used to derive the general formula shown. In the same way power costs when they become a major factor can become also a part of the "*BC* Factor."

Quite obviously, future economic conditions may dictate the calculation of a new family of curves showing the return. The general formula, with constants to suit the changed conditions, would still apply for the approximate solution of the general problem.

Discussion

L. B. Curtis (Pennsylvania Railroad Company, Philadelphia, Pa.): In this excellent paper Messrs. Brown and Marmaros have made a valuable contribution toward the future of electrification of railroads in this country. The curves in Fig. 1 should prove a very helpful tool to management in determining whether a detailed study of electrification is warranted.

The authors under the heading "The Difference in Investment" state, "For electric operation, the investment involved consists of: 1. The cost of the contact system (a function of *A*)." In paragraph *A* is the statement "The miles of track to be considered as electrified to handle the required traffic. This determines the additional investment required for the contact line and any other costs related thereto, etc." This broad statement probably includes the cost of providing proper electrical clearances for the 25-kv system. In some areas this item might be negligible whereas in others the cost might be very high. In England it was found necessary to reduce the contact voltage by 75% in many of the restricted overhead areas. With railroads being crossed more and more by overhead highway bridges, and with tunnel and station overhead clearance restrictions this clearance cost could be an important factor in the cost of the contact system. Does the value K_1 include an amount for clearances, and if so how much?

The authors also state, "2. Higher speeds would reduce the investment in car equipment, and motive power although this might require more electric horsepower than diesel in some cases." It would be helpful if the authors would explain this statement in more detail. It is my understanding that everything else being equal, the higher the speed of operation, the more justified economically is electrification.

J. Stair, Jr. (Gibbs & Hill, Inc., New York, N.Y.): This is a universally applicable approach to determine whether a certain section of railway warrants a detailed study to ascertain whether its electrification will result in more economical operation. The economy of electrification depends, in measure, on the traffic density, which this paper takes into account.

At the 16th Session of the International Railway Congress Association held in London in 1954, the French National Railways reported: "Except for lines for which specific technical reasons arise, the programs of electrification depend entirely upon considerations of an economic order. The practicable principle adopted is the consumption of coal. The SNCF (French National Railways) feels that electrification with 1,500 volts direct current is convenient when the annual consumption of coal per kilometre of double track exceeds 400–500 tons, whilst for electrification with standard frequency single-phase current this limit is reduced to about 300 tons."

Of course, these French figures cannot be applied universally; indeed, with changing economic conditions, the French may modify them from time to time.

G. J. Sennhauser (Chesapeake and Ohio Railway Company, Cleveland, Ohio): The authors are to be commended for their unique and simple method for the investigation of the economic justification of electrification. However, it would seem that the authors are a bit optimistic in anticipating the rate of return on the investment necessary to effect electrification.

The Chesapeake and Ohio Railway's operations, on the mainline from Toledo, Ohio, to Newport News, Virginia, entitle it to occupy a position on Fig. 1 that would indicate a rate of return, on the investment to effect electrification, of over 20%. This, as indicated by the paper, is deserving of a more thorough and searching study. Such a study has been made in conjunction with one of the country's largest electrical manufacturers and the anticipated rate of return in 1962 (first full year of electric operation) was not the indicated 20+% but less than 5%! A rate of return of 5% or less on a project of this nature makes it virtually impossible to attract outside financing and no railroad would seriously consider financing the project itself.

What are the factors that bring about this difference of opinion? I believe them to be:

An unrealistic comparison of the costs of owning and operating diesel electric locomotives versus straight electric locomotives. The authors estimate a straight electric locomotive to cost 6.81¢ per thousand rail hp unit per mile, and a diesel electric locomotive to cost 37.82¢ per thousand rail hp unit per mile. This indicates a cost ratio of 5.55 to 1. These costs may be realistic on the basis of industry averages. However, they certainly cannot be construed to be realistic on the basis of identical unit age, service and utilization. It is doubtful that more than a 2 to 1 cost ratio could exist on a unit mileage basis in favor of the straight electric locomotive in identical service and the true operating costs would be tempered by maintenance to the overhead contact system.

The authors, in their study, allow $30,000 per track mile for the cost of the overhead contact system, $10,000 per track mile for switching and control system and $10,000 per track mile for changes to the signal and communication systems. The estimates for making the necessary changes to the signal and communication system on the Chesapeake & Ohio amount to more than $35,000 per track mile! It is reasonable to assume that any railroad having a traffic density great enough to warrant considering electrification is equipped with a very modern signal and communication system and that the costs of making it compatible with electric operation will be considerably in excess of $10,000 per track mile assumed by the authors.

In the past, trunk line railroads in the United States have been electrified because of heavy grades, smoke problems in tunnels, smoke abatement in cities and terminals, and high traffic density; especially where a large number of passenger trains are run and the acceleration characteristics of electric traction may be used to good advantage. However, today, centralized traffic control, the diesel-electric locomotive, train radio, and other technological advances allow the railroads to cope with operating adversities of this nature without resorting to electrification—indeed in some instances, electric operation has been displaced by the diesel. Future electrification must justify itself on the basis of operating economy alone and must be able to offer a true return of at least 20% on investment. The present state of the electrification art coupled with the prevailing price relationship between electric power and liquid fuel will not induce the railroads to scrap their diesels and commence stringing wire. Technological progress and changes in management philosophy will succeed in reducing the capital requirements for electrification. However, one factor will determine whether or not the railroads electrify. This is the cost relationship between liquid fuel and electric energy. Only a significant improvement in favor of electric power will cause the railroads to turn to electric operation.

H. F. Brown and H. E. Marmaros: Before answering in detail the various questions raised by those who have discussed this paper, attention is directed again to one word in the title. This paper was purposely intended to be an *approximate* method

only of determining whether railroad electrification applied to a given route and traffic density can indicate a return attractive enough to warrant a more detailed and serious study. Such studies usually entail expense if they are to become the basis for substantial capital expenditures.

The costs on which the curves shown in Fig. 1 are based were purposely omitted from the paper. They fall into two general categories: 1. those dealing with operation, which are now fairly well known or which can be determined with a fair degree of accuracy from recorded statistics; and 2. those dealing with new investment, some of which are known and some of which must be calculated or estimated by engineering knowledge and experience in this field.

The first group of costs are hardly debatable except possibly as to future trends. Some of the second group pertaining to investment can be a matter of judgment and opinion.

By omitting all costs and by presenting a simple method whereby an approximation of the economic value of electrification could be easily determined by any interested railway, it was believed that the paper would have wider interest and application.

The cost of new construction can vary over a wide range for different railroad operations. Fairly dense freight traffic with average passenger service and no commuter service was envisioned in this paper as the general field for future electrification. The investment costs selected were based on round figure estimates for such operations.

It should be stated here that several of those who have discussed the paper asked for, and were given, all the general costs and values of the constants used by the authors in developing the formula and the curves shown in Fig. 1.

Mr. Stair's comments are appreciated, coming from one who has made valuable contributions to railway electrification during an outstanding career in this field.

He mentions the formula developed by the French, based on fuel consumption per track mile (or kilometer) to determine their railway electrification economy. This country has been so abundantly blessed with large supplies of relatively cheap fossil fuel, both solid and liquid, that fuel saving has never been considered as a motive for railway electrification in the United States. The day may not be far distant, however, when this factor will have to be considered in our railway operations, especially in connection with liquid fuel. Fuel saving, per track mile, is another function of traffic density.

In a recently published article,[1] the financial return on the most recently installed 50-cycle 25-kv electrification of the French National Railway System is stated to be 10.3%.

The greater fuel saving realized with this latest system as compared with previous installations at lower voltage, is of interest. The use of this high voltage commercial frequency system is one of the basic assumptions in this paper.

Mr. Curtis asks about the costs of providing clearances for the 25-kv contact system, and whether such costs were specifically included in the formula given.

As stated in the foregoing, only round-figure estimated costs were assumed for the new contact system. No specific costs were assumed for greater clearance than have been required on the existing 11-kv installations in this country. The installations of the 25-kv 50-cycle system used in France, Portugal, Turkey, and Belgian Congo, have insulation and clearances quite comparable to those which have been used in the several 11-kv installations here.

If especially limited clearances exist on a railroad for which electrification appears to have economic possibilities, the cost of providing additional clearance would have to be determined, with other special costs, by any detailed study which might be warranted.

Mr. Curtis asks further for a little more elaboration of the statement made in the paper that:

"Higher speeds would reduce the investment in car equipment and motive power, although this might require more electric horsepower than diesel in some cases."

The formula used in this paper assumes that diesel motive power is replaced with electric motive power having the same total equivalent rail horsepower. This would show no increase in speed. However, the short-time rating of the motors of electric locomotives can be used more effectively than with diesels, and also more horsepower can be built into an electric locomotive unit than into a diesel unit. These higher horsepower units can add up to more total horsepower than with diesel units. The increased horsepower available can produce higher speeds, which can haul more tonnage per hour or the same tonnage in a shorter time. A faster schedule could reduce the time of use of car equipment per run in such service as hauling coal or ore from source to destination in intensive repetitive operations. Thus, more trips per week could be made, and fewer cars, and possibly locomotives, would be required.

Mr. Sennhauser's discussion is indeed challenging because of the position on the chart shown in Fig. 1 the railroad with which he is identified might occupy.

The authors are not familiar with the details of the "more thorough and searching study" to which he refers and cannot, therefore, comment on the factors used which show such a low rate of return as 5%.

It is the opinion of the authors that if there is any one region in this country where railway electrification should be able to prove its economies, it is on the railroads in the Pocahontas region, especially those railroads which are hauling coal traffic of high density. One such railroad has been partially electrified for a number of years and the low maintenance cost of its motor-generator locomotives is the subject of a valuable unpublished paper by T. F. Perkinson.

This region is well supplied with electric power, which, it is understood, is available to the railroads today at rates which are on a par with the present costs of liquid fuel in this area.

Mr. Sennhauser believes the authors have been unrealistic in their "comparison of the costs of owning and operating diesel electric locomotives versus straight electric locomotives . . . on the basis of identical unit age, service and utilization"; and expresses doubt "that more than a 2 to 1 cost ratio could exist on a unit mileage basis in favor of the straight electric locomotive . . ."

Since these cost differentials are the crux of the economies which can be made with electric operation, the authors welcome the opportunity to show just how they have been derived.

These costs are all in the first category mentioned in the second paragraph of this closure. They are not estimates. They are known costs, carefully compiled from recorded operating statistics covering both types of motive power. They are not costs "on the basis of identical unit age." They are the *average annual costs* per thousand rail hour power mile *over the economic life of the equipment*. The economic life has very carefully and conservatively calculated from the "rate of rise" in these maintenance costs, to be 15 years for the diesel and 25 years for the electric locomotive.

Thus, the figure 6.81¢ per 1,000-rhp unit mile for electric locomotives represents a cost to which maintenance will rise between the 12th and 13th years, this being the average for a 25-year economic life. Comparison is invited between this figure and the cost of maintenance shown in Mr. Perkinson's paper for electric locomotives now 9 years old. Indeed, the Virginian locomotive costs are proving to be even lower than the 6.81¢ figure used in this paper. These locomotives are in a service quite comparable to that of the diesel locomotives on the railroad with which Mr. Sennhauser is identified.

Next, take the cost of 37.82¢ per 1,000 rhp mile for diesel locomotives. This represents the cost to which repairs will rise between the 7th and 8th years, this being the average for a 15 year economic life. This figure, and the economic life of 15 years was determined from a careful analysis made by the authors of the repair costs of several thousand diesel locomotives related to their age on a number of railroads. This investigation included the diesel locomotives on the railroad with which Mr. Sennhauser is associated. He doubtless knows how realistic this investigation had to be.

Some of the manufacturers and most of the users of diesel locomotives have, to date, not realized, or have been unwilling to admit the high rate of rise in maintenance costs, and the relatively short economic life of this type of equipment. This is because the equipment, on the average, is still too new and has been acquired over too short a period to clearly show these trends. The "average age" of all this equipment in use has been, until 1954, relatively stationary between 2 and 3 years. Now that the railroads are almost completely equipped with this type of motive power, its average age, and maintenance costs will rise more rapidly during the next few years.

Mr. Sennhauser can be assured that this differential is much greater than the 2 to 1 ratio he assumes. Even Mr. Perkinson, a well-known and conservative authority, says in his paper today, with reference to the repair costs of the electric locomotive under discussion, that these costs "represent . . . . . a value approximately 1/4 to 1/3 of comparable diesel-electric locomotive maintenance costs for the same type of service performed."

In 1956, all road diesels on Class I railways, approximately 17,500 in number, cost $365,000,000 for repairs. This is $20,900 per unit for diesel units averaging 6 years

in age, and for all road services both passenger and freight. Possibly Mr. Sennhauser will agree that the costs of maintaining diesel locomotives on his railroad should be slightly higher than the national average for Class I railroads on the basis of the unusually intensive service they perform.

Maintenance costs, in spite of the substitution of diesel power for steam still represent the largest single item in the total cost of operating railroad motive power on the Class I railroads. Only in the post-war period 1945 to 1950, when fuel costs were abnormal, was the cost of fuel higher than the cost of maintenance. In 1956 maintenance represented 29.1%, fuel 26.3%, and wages of enginemen 25.5% of the total costs of road locomotive operations on the Class I railroads.

Mr. Sennhauser rightly calls attention to the high costs which can be incurred for signal and communication changes necessary for electrification. These costs fall within the second category mentioned above.

The authors allowed a round sum of $50,000 per mile for the contact system, power control and changes to the signal and communication system. Contact line and power control estimates have been made in some instances as low as half of this amount, leaving the other half or $25,000 for signal and communication changes.

It is admitted that a careful study for a particular railroad might show all these costs to be higher or lower. However, the formula is given and the paper clearly states that other curves can be developed to meet other conditions than those assumed by the authors. For example it can be calculated that if the contact line and power control should cost $40,000 per track mile, and that signal and communication changes cost an equal amount, making a total of $80,000 per mile, the formula given would indicate a new family of curves with the point P_3 on Fig. 1 still showing a return of slightly over 10% (which is considered attractive in Europe).

If signal and communication systems have to be drastically revised because of electrification, the entire cost of the new plant is not necessarily a charge against electrification. Only the unamortized value of the old plant, and the differential in cost of the new plant over the old plant would be the proper charge.

The railroads are spending large sums annually for revisions to their signal and communication systems. Some of these systems cost as much per mile of track as electrification.

One way to keep the cost of future changes down would be to first examine whether traffic indicates that electrification might have economic possibilities, by the simple method outlined in this paper. Then, since it may take a few more years to wear out the present motive power and to perfect plans, start now to make all future signal and communication changes so that they can be made compatible, with the least possible future expense, with electrification when it comes. The authors believe that if this is done, the costs of such changes can be kept within the limits assumed. On this basis, there is no point in overloading the costs for the first trial assay.

Finally, Mr. Sennhauser expresses the opinion that the only factor which will cause railroads to electrify will be the "cost relationship between liquid fuel and electric energy."

Again he is referred to the relation between maintenance costs and fuel costs mentioned previously. Both of these costs have been of the same relative magnitude during the years 1940 to 1950, during which period the diesel has superseded the steam locomotive. It seems obvious that even a 2 to 1 reduction in maintenance costs will produce as much economy as a 50% rise in liquid fuel costs over present electric energy costs. Where this ratio in reduced maintenance costs is between 3 and 4 to 1, as Mr. Perkinson indicates, or even more, as the authors believe, the operating savings which can be realized right now are as great as can be realized from a 75 to 90% differential in fuel-electric energy costs at some indefinite time in the future.

The authors are not advocating the scrapping of any usable motive power, or the stringing of any wires at this time. They have been aware for several years of the large differential in the maintenance costs of the two types of motive power. They are also aware of the railroads' growing necessity for more economic operation.

Present motive power will have to be replaced within a few years. Purchased electric power at attractive rates is becoming more and more available to the railroads. The authors have felt that there are many places where traffic density would warrant giving electrification a serious study. In this paper they say, "Here is a simple measuring device. Take it and measure your railroad, and see if electrification can save you some real money. If so, take a closer look and plan ahead accordingly."

The authors are grateful for the discussion, as it has given opportunity for presenting additional supporting data for which there was not space in the paper.

Reference

1. Electric Traction on the Railways. *Bulletin*, International Railway Congress Association, Brussels, Belgium, Sept. 1957.

Superposition Applied to Mechanical Rectifier Contacts

I. K. DORTORT
MEMBER AIEE

CONTACTS of mechanical rectifiers physically make and break connections between the a-c phase terminals and d-c terminals of the rectifier, but do not contribute to the important phenomena of commutation. Commutating reactors in series with the contacts are used to produce the required conditions during "make" and "break" steps in which the contacts are able to perform their limited functions. The main power circuits may be seen in heavy lines in Fig. 1. Make and break phenomena are equally important, but this paper deals only with break problems, the make problems having been adequately solved in modern rectifiers.

Rectifier contacts are not in any sense high-speed circuit breakers, since no arcing can be tolerated. An arc lasting only a fraction of a millisecond can carry over into a short circuit, commonly called a "backfire." Moreover, at 5 million operations a day, exceedingly small currents and voltages can, without visible arcing, produce intolerable material transfer and loss of timing, resulting in backfire. To assure reasonable contact life and stable operation, the following minimum requirements must be met:

1. Breaking conditions must be nearly perfect over at least the middle third of the break-step.

2. Within this normal operating range of the step, the contact current at the instant of break must not exceed 0.5 ampere. In a typically rated 5,000-ampere double-way rectifier, this represents 0.01% of the peak contact current.

3. At the instant of break, the total voltage across the contact including any inductive voltages must not exceed 10 volts. In a 400-volt rectifier this represents about 2% of the peak commutating voltage.

Obviously, if the current and voltage are smaller than the limits just given, contact life will be increased. By the

Paper 58-137, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted September 18, 1957; made available for printing November 12, 1957.

I. K. Dortort is with the I-T-E Circuit Breaker Company, Philadelphia, Pa.

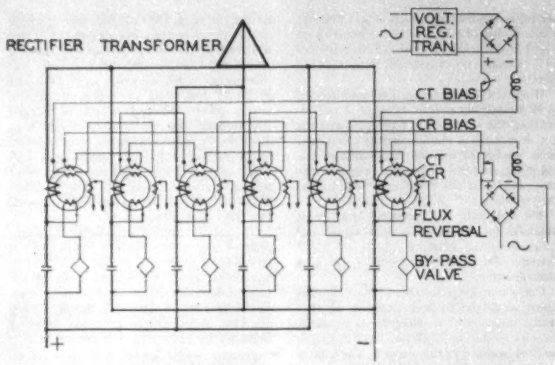


Fig. 1. Diagram of 6-coil 6-phase double-way rectifier, incorporating combined CR (commutating reactor) and CT (current transformer), and bias for CR

same token, if these limits can be maintained over a wider part of the break-step, operation of the rectifier will be more stable.

In earlier mechanical rectifiers, the imperfections of commutating reactors were compensated by complex excitation circuitry which could provide accurate correction over a limited range of operating conditions and a relatively small part of the break-step. Adjustments were generally difficult, time-consuming, and not very stable. The new circuit described in this paper overcomes the difficulties and limitations simply by eliminating the excitation circuits and substituting for them a bypass circuit which superposes on the contact current an oppositely directed current, inherently equal at every instant to the break-step current. Before the theory and operation of the new circuit are introduced, a brief description of the function of the commutating reactor and conventional circuitry is in order.

Conventional Excitation of Commutating Reactors

The principal function of the commutating reactor is to extend the natural current-zero of an a-c wave in order to provide a finite period of time during which the contact may be opened safely. The rate of decrease of current just prior to contact separation in large rectifiers is often 5 million amperes per second or more. In the absence of commutating reactors, missing the exact current-zero by even one microsecond would cause the contact to break about 5 amperes, or 10 times more than permissible for reasonable life expectancy. Considering that current-zero may vary unpredictably in each cycle due to changes in voltage, load, wave-shape distortion, and other factors, it is easily understood why the realization of the mechanical rectifier awaited the development of the commutating reactor.

Even if it were possible to predict exact current-zero, and cause contact separation at exactly current-zero, the full commutating voltage existing in the circuit would, in the absence of the commutating reactor, appear instantly across the opening contact.

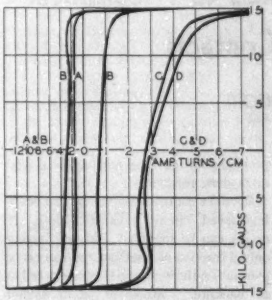


Fig. 2. Typical magnetization curves

A—Static characteristic
B—Dynamic, at 25 kilogausses per millisecond
C—Same as B except on expanded scale
D—40 kilogausses per millisecond, on expanded scale

The ideal commutating reactors would have zero-magnetizing current up to a sharply defined saturation point beyond which there is no further flux change. The nearest approach is a reactor with a relatively narrow, rectangular, hysteresis loop. The core material most commonly used is a 50-50 nickel-iron alloy, cold-rolled to 2 mils to develop its full magnetic properties and to reduce eddy current losses to a minimum. This material has a coercive force of approximately 0.2 ampere-turn per centimeter and a B_{max} of 15.5 kilogausses.

A typical magnetization curve is shown in Fig. 2. Curve *A* is the static characteristic. Curve *B* was taken at 25 kilogausses per millisecond. The horizontal scale was chosen to display the complete loop in the proportions generally used.

Curve *B* plotted on a compressed horizontal scale gives the impression of being quite straight. The same curve plotted on an expanded scale in *C* shows up the irregularities which occur and looks much less favorable. Curve *D* plotted on the same expanded scale was taken at 40 kilogausses per millisecond. All curves were taken at uniform speeds of magnetization and do not represent true operating conditions, since the impressed voltage varies throughout the step.

Because of losses, reactance, and cost, it is not practical to build commutating reactors for large rectifiers with a magnetizing current of less than 2 amperes. Compensation by pure d-c bias can reduce the step current through the contact but is not sufficient in itself to provide satisfactory matching over the entire

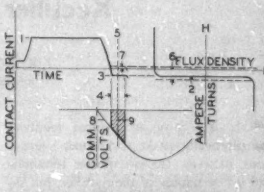


Fig. 3. Behavior of commutating reactor during break-step

1—Peak power current in contact
2—Magnetization characteristic
3—Uncorrected break-step current
4—Duration of break-step
5—Point of contact separation
6—Reactor bias
7—Compensated break-step current
8—Commutating voltage
9—Reactor volt-seconds

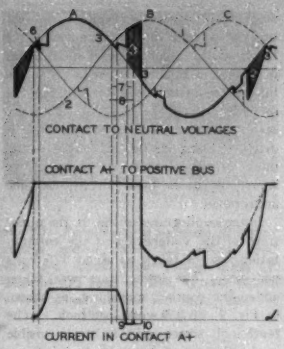


Fig. 4. Rectifier relationships

1—Positive voltage envelope
2—Negative voltage envelope
3—Voltage, contact A+ to neutral
4—Break-step voltage
5—Flux reversal voltage
6—Make-step voltage
7—Phase current overlap
8—Mechanical overlap
9—Uncorrected break-step current
10—Completion of break-step current through bypass circuit

available step under widely varying conditions.

The upper left-hand part of Fig. 3 shows the contact current in idealized form, with the make and break step currents exaggerated for clarity. Without external excitation applied to the commutating reactor, the break-step current is negative with respect to the power current flow. In line with this current diagram, and lying on its side, is the magnetization curve of the reactor represented as being perfectly straight. If a d-c bias is applied to the commutating reactor, exactly equal in ampere-turns to the magnetization characteristic, there will be no current flowing in the main winding of the reactor during the break-step, and the current through the contact will be zero at that time. Because of the irregularities in the characteristic and because of the variations imposed by operating conditions, the bias is generally made somewhat greater than required, as shown. The reactor is overcompensated and a small positive current will flow through the contact.

Just below and aligned with the current trace of Fig. 3 is seen the commutating voltage and the portion of this voltage wave absorbed by the commutating reactor. Very little of the commutating voltage appears across the

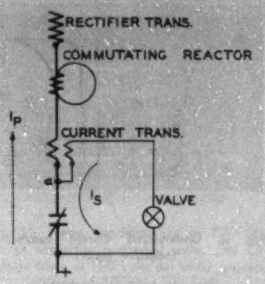


Fig. 5. Functional diagram of superposition circuit

I_P—Primary current of CT (magnetizing current of commutating reactor)
I_S—Secondary current of CT, net contact current (forward) $I_S = I_P$

contact on opening, provided the residual contact current is shunted around the contact through a low impedance bypass circuit. Note that the voltage across the commutating reactor is increasing during the break-step, thereby changing the speed of magnetization of the reactor and causing additional deviation from the magnetization curves of Fig. 2. The position and therefore the length of the break-step varies widely due to changes of alternating voltage, load, phase control, and other causes. The composite alternating voltages, contact voltage, and contact current of a 6-phase double-way 6-coil rectifier are given in Fig. 4.

To extend as far as possible the useful part of the break-step, "straightener circuits" are employed to compensate for the irregularities of the hysteresis loop. Unless their contribution is kept small these circuits not only add to the over-all complexity, but are very difficult to adjust. Their action also varies with operating conditions and therefore requires additional adjustment of the excitation.

Over and above the special excitation and straightener circuits, bypass circuits around the contacts are important, to hold down the voltage after the break and to reduce the effect of unavoidable voltage spurs at the instant of contact separation. These voltage spurs are indicative of stored energy which is released at the contact and dissipated in minute sparks.

Because so many variables affect the magnetization characteristic, it is usually necessary to readjust the excitation with varying operating conditions. In many installations this is done manually. When wide ranges of automatic regulation are specified, it is necessary to vary the excitation automatically in response to varying conditions. Since the adjustment required is not linear and differs for each of the variables, the circuitry becomes exceedingly complex.

All in all, the special circuitry and equipment required to maintain "black commutation" by conventional methods, is complex, bulky, and expensive. Constant monitoring and adjustment are required to maintain optimum breaking conditions under widely varying conditions of supply voltage, output voltage, load, and temperature.

Cancellation by Superposition

To reduce the current through a contact to zero, all that is needed is to superpose an opposite current of exactly the same instantaneous magnitude. A perfect current transformer, operative only during the break-step, and a current-limiting valve to prevent excessive power losses, are the means used to accomplish this. Fig. 5 is a functional diagram of the basic circuit.

Current Transformer

The first requirement is a perfect current transformer of 1-to-1 ratio, capable of carrying, in its primary only, the full load current of the contact, and reproducing accurately in its secondary the magnetizing current of the commutating reactor during the break-step.

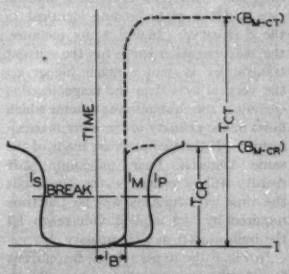


Fig. 6. Operation of current transformer

I_P—Primary current
I_M—Magnetizing current
I_B—CT bias
I_S—Secondary current
T_{CR}—Length of break-step to saturation of CR (B_{M-CR})
T_{CT}—Step-length of CT to saturation (B_{M-CT}) at bypass circuit voltage

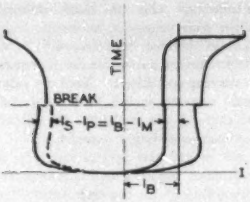


Fig. 7. Overbiased current transformer (positive break)

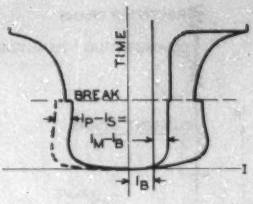


Fig. 8. Underbiased current transformer (negative break)

The same magnetic material used in the commutating reactor is suitable for the current transformer. Assuming a current-limiting valve with practically zero voltage drop when gated, the burden on the current transformer and therefore the rate of magnetization of the core will be small: by making the outside and inside diameters of the core nearly equal, the hysteresis loop will have nearly rectangular sides. The rising magnetization curve of the current transformer may then be almost perfectly cancelled by a pure d-c bias applied through an auxiliary winding. The primary and secondary currents must perforce be equal as long as the core is unsaturated.

The principle of operation of a current transformer of this type in a rectifier circuit is illustrated in Fig. 6. Currents are plotted along the horizontal axis, with time along the vertical, in order to emphasize the similarity of these curves to the conventional *B-H* curves. When volt-seconds are applied at a uniform rate, flux density increases at a uniform rate and the plots become identical to the *B-H* curve. In Fig. 6, for instance, the magnetization curve for the current transformer is shown much higher on the vertical axis than the magnetization current of the commutating reactor which flows in the primary of the current transformer. If the two cores are made of the same material their maximum flux density will of course be the same, but the time to saturation will be the time required by the applied voltage to fill the built-in volt-second capacity.

In Fig. 6 the upper part of the current transformer magnetization curve is shown in dashed line. Until the commutating reactor reaches saturation, the voltage across the current transformer is very low and quite uniform. As the reactor-magnetizing current approaches saturation value, the voltage across the current transformer increases rapidly, so that saturation of both cores occurs almost simultaneously.

In the first example a bias just equal to the magnetizing ampere-turns is applied to the current transformer and the secondary current is therefore an exact image of the primary current. Oppositely superposed on itself in the contact, the current in effect commits suicide. Since the contact current is now zero, opening of the contact does not change anything, and the break-step of the commutating reactor is completed through the bypass circuit. Whether the magnetizing current of the current transformer is larger or smaller than that of the commutating reactor is immaterial as long as the required ampere-turns to match it are supplied by the bias.

In Fig. 7 we see a similar condition except that the bias is slightly greater than the magnetizing current. The secondary current of the current transformer is greater than the primary current by the same amount that the bias exceeds the magnetizing current. Now, a slight change does occur at the instant of contact separation since beyond this point the two currents must again be exactly equal. A slight redistribution of voltages occurs, producing small breaks in the magnetization curves, as shown.

The bias may be made slightly smaller instead of larger than the magnetizing current. In this case the residual contact current is negative instead of positive. A similar phenomenon takes place at the break but in the reverse sense. This mode of operation illustrated in Fig. 8 is preferable to the larger bias if the cross section of the current transformer core is large. With excess bias, most of the unexpended volt-seconds of the core appear across the contact immediately after contact separation, rather than after the break-step.

The principal reason for mismatching the bias is to obtain the small discontinuity in the magnetizing current, so that the position of the break can be observed on a high resolution oscilloscope. To locate the beginning of the break-step and further accentuate the break on the oscilloscope, a small shunt is inserted in the bypass circuit, as shown in Fig. 9. Other connections for the oscilloscope may be used to locate the break: this one was chosen because it displays other useful information and most nearly resembles the traces to which operators have already become accustomed.

In the oscillograms of Fig. 10 the upper trace is the voltage across the commutating reactor taken at about 175 volts per inch. The lower trace was taken across the contact and shunt, as shown in Fig. 9, at 350 volts per inch. (The tenth-inch screen grid is barely visible in the photograph.)

These shots were repeated in Fig. 11 at approximately 100 volts and 10 volts per inch respectively, and amplified on the horizontal scale also. The break cannot be seen on the commutating reactor trace. It shows up on the shunt and contact voltage trace at slightly more than 1/2 volt. Only half of this appears across the contact. Because the recovery voltage across the contact is so small, the residual current may be safely adjusted to zero, permitting small excursions in either direction.

It is readily apparent that the commutating reactor magnetizing current determines the rating and physical size of the valve components. A d-c bias applied to the commutating reactor can reduce this current appreciably. Exact adjustment and regulation of this bias is not required, and considerable advantage is gained at little cost.

A single-turn through-type current transformer is most economical in dollars

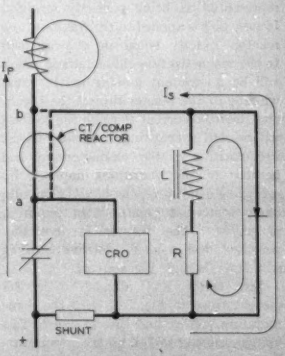


Fig. 9. Simplified version of bypass valve

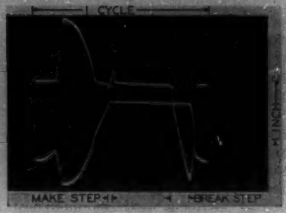


Fig. 10. Performance at maximum alternating voltage and maximum flux reversal, 330 volts d-c. Upper trace—voltage across CR taken at 175 volts per inch. Lower trace—voltage across contact and shunt, taken at 350 volts per inch

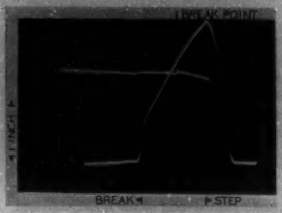


Fig. 11. Same conditions as in Fig. 10, except expanded horizontal scale, and CR voltage taken at 100 volts per inch, contact voltage at 10 volts per inch

as well as space requirements and losses. The circuit configuration of Fig. 9 is essentially the same as that of Fig. 5 except for the 1-turn current transformer and details of the current valve. Note, however, that the secondary circuit need not be connected to the primary at point a as shown in dashed lines in Fig. 9, but is connected at point b as shown in solid line. Obviously there is no electrical or magnetic difference. However, with the connection made at b the configuration of the circuit looks different even though no change has been made. The current transformer now looks like a nonlinear reactor and, in fact, the circuit operation may be quite readily explained from this point of view. For this reason the current transformer is frequently called the "compensating reactor." The same connection and the same logic apply if the compensating reactor has more than one turn, the multiturn reactor then being the exact equivalent of a multiturn 1-to-1 ratio current transformer. The current through the reactor, and therefore through the contact, is $I_B n_B/N - I_M$, in which I_B and n_B refer to the bias winding, and I_M and N refer to the main winding of the reactor.

If the number of primary turns on the current transformer is made the same as that on the commutating reactor, these turns may be made common to both. The cross section of the current transformer core is less than 1.0% of that of the commutating reactor, so that the physical size of the combined reactor and current transformer is very little different from that of the reactor alone. There is practically no increase in copper losses and air-core reactance, so the efficiency and the regulation of the rectifier are not significantly affected; see Fig. 12.

Valve

The second important requirement for this circuit is a valve which must present a very low impedance to the bypass current during the break-step in order to assure a low recovery voltage across the contact and to hold down the size of the current transformer and associated components. This same valve must present, in the same direction, a high impedance immediately after the break step is completed. The inverse voltage produces currents through the valve in the same direction as the bypass current just preceding the inverse voltage period. If the impedance of the valve is not high during the inverse period, the losses will be great, affecting the overall efficiency of the rectifier.

Fig. 9 is a simplified version of the valve actually used in somewhat more complex form for this application. During the inverse voltage period, current flows from the positive bus (in this case) through the resistor and choke coil, to the then negative terminal of the rectifier transformer. When the contact closes, the voltage across this circuit disappears, but the choke coil keeps current flowing in the loop composed of the resistor, choke coil, and contact. When the break-step begins, all of the reactor discharge current is transferred to the diode branch, since the compensating reactor is now unsaturated. The bypass current also flows through the diode, in the reverse direction, superposed on the reactor discharge forward current. The net diode current must, of course, be in the forward direction. In other words, the time constant of the reactor discharge circuit must be long enough to assure sufficient current during the break-step to carry the counterimposed bypass current without blocking the diode during the break-step.

Note that the voltage drop across the single diode is opposite in sense with respect to the bypass current and the small resistance in series with the valve is required to produce a voltage in the right direction across the compensating reactor or current transformer. If the resistance drop equals the diode forward voltage, the net drop is zero. This resistor is also used as a shunt to make the break observable on the oscilloscope.

Fig. 13 shows another suitable valve, similar to the Koppleman bypass valve. During the inverse voltage period, current flows from the positive bus (in this case) through diode 1, the resistor and choke coil, and diode 2, to the now negative terminal of the rectifier transformer. When the contact closes, the voltage across this circuit disappears, but the choke coil keeps current flowing through two parallel branches consisting of diodes numbered 1 and 3, and 2 and 4, with no current in the external leads. The time constant must be such that sufficient current is available in diodes 3 and 4 to carry the counter imposed bypass current without blocking throughout the break-step. The voltage drop across the valve, and therefore the recovery voltage across the contact at the time of opening, is the difference of the forward drops of diodes 1 and 4 or diodes 2 and 3 at their respective currents during the break-step. Since these cell drops are small to start with, the difference voltage is almost negligible.

Although this circuit has been tried and found to operate satisfactorily, the choke coils required to obtain the proper time constant are excessively large. The time constant may be reduced by using low values of current-limiting resistance with a resulting large increase in losses, since the losses are inversely proportional to this resistance. Even so, the circuit is attractive when the alternating voltage impressed on the mechanical rectifier is

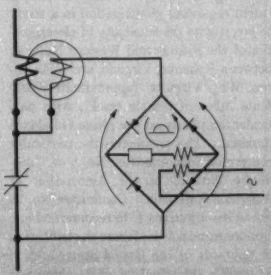


Fig. 12. Modified bypass valve with external power source to supply half-wave current during gating: 1/1 CT built into CR

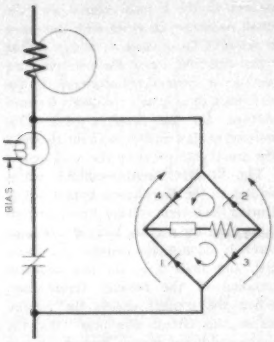


Fig. 13. Full-wave valve and biased compensating core

maintained approximately constant. When, however, voltage regulation is obtained by decreasing the alternating voltage, the inductance-resistance circuit must be proportioned for the lowest voltage at which it must operate, and losses at maximum voltage then become prohibitive.

For these reasons a variation of this circuit was devised; it is illustrated in Fig. 12. Here the choke coil is replaced by a small transformer or by individual windings of a multiphase, multiwinding transformer, energized at constant voltage from the auxiliary power source. The voltage applied to the small auxiliary transformers is so phased that the resulting half-wave current covers the maximum ranging of the break-step and, during that portion of the cycle, it is always larger than the maximum bypass current. The current contributed by this transformer during the dwell time of the contact is independent of the voltage applied to the rectifier. The loss due to the inverse voltage of the rectifier varies as the square of this voltage. Careful proportioning of all circuit components is required to hold losses to a minimum.

Conclusions

Fig. 1 illustrates the relative simplicity of the new excitation and bypass system applied to a 6-coil 6-phase double-way rectifier, including d-c bias for the commutating reactors, but excluding details of the valve circuit, which are shown in Fig. 12.

Rectifiers equipped with this circuit have now been in commercial use more than half a year. Their performance is eminently satisfactory and operation incomparably simpler than with previous rectifiers. In one case in particular, voltage control was provided over a range of 50% by means of an alternating voltage regulator and an additional range down to approximately 20% by means of magnetic phase control. The full range of load and voltage control is accomplished without disturbing initial adjustments. In another installation wide swings of primary voltage on a poorly regulated high-voltage line are accommodated without ever touching the excitation circuit adjustments. In both cases the rectifier contacts seem to be standing up very well, although the relatively short period of operation does not yet permit a true evaluation of contact life with the new system.

Reference

1. Commutating Reactor Control for Mechanical Rectifiers, Edward J. Diebold. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 1062–65.

Technical Aspects of Providing Service to Single-Phase 60-Cycle Railroad Loads

T. J. NAGEL
MEMBER AIEE

A. F. GABRIELLE
ASSOCIATE MEMBER AIEE

IN EARLY 1956, representatives of Gibbs & Hill, Inc., approached the American Gas and Electric Service Corporation regarding participation in a study to investigate the feasibility of electrification of the Norfolk and Western Railway between Roanoke, Virginia, and Williamson, West Virginia, (approximately 200 route miles of double track), with particular emphasis on the phase unbalance problem. This problem is the particular concern of Subcommittee No. 5 of Committee No. 13 of the Association of American Railroads. Committee No. 13 has as its objectives 1. to recommend one or more standards for future electrification of railroads in the United States and, 2. to resolve the effect of single-phase unbalance imposed on 3-phase utility systems by single-phase loads from 60-cycle electrified railroads.

Following discussions with Gibbs & Hill engineers, six points of power supply, advantageous to both the Norfolk and Western Railway and the American Gas and Electric Company system, were chosen. The necessary load information, furnished by Gibbs & Hill, was based on a representative day of heavy traffic movement consisting mainly of 20 freight trains moving eastward at an average of one every 72 minutes with a load of 5,200 to 7,000 tons.

Problems Associated with Single-Phase Loads

A major problem associated with the supply of single-phase power from a 3-phase system is the resultant heating effect encountered in turbogenerator rotors.

Single-phase and 3-phase instantaneous power differ in that the former is a pulsating quantity while the latter is of constant magnitude. Therefore, to serve a single-phase load from a 3-phase 60-cycle generator requires that the machine store and release energy at a rate of 120 cycles per second. Hence a generator rotor must carry 120-cycle current (commonly referred to as the negative-sequence component) which flows partially in the field windings and partially in the surface steel of the rotor. The rotor surface currents flow axially toward the ends of the rotor and then cross the slot wedges where the heating is usually most severe. The temperatures developed are unknown at the present time and the problem is complicated by the random nature of contact resistance phenomena (location of wedges, condition of surface, centrifugal force, etc.).

The present operating recommenda-

Paper 57-1096, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted June 10, 1957; made available for printing January 2, 1958.

T. J. Nagel and A. F. Gabrielle are with the American Gas and Electric Service Corporation, New York, N. Y.

tions of the General Electric Company for unbalanced loading of turbine-generators are given in Fig. 1.[4] Operating practices generally have been such as to minimize unbalanced loading because of the hazard to generators.

In view of the need for additional data pertaining to permissible generator current unbalance, a joint program has been undertaken by the General Electric Company and the American Gas and Electric Service Corporation, which consists of the following:

1. Development and successful application of continuous temperature measuring devices for use during factory testing of large rotating fields. The General Electric Company has been working intensively on this project during the past several years and techniques gradually are being evolved for the use of thermistors and thermocouples for continuous temperature measurement.

2. Measurement of rotor iron temperatures on several large fields during factory unbalanced loading tests. These tests will provide more up-to-date data on the temperatures to be expected in present-day design.

3. Laboratory tests on simulated wedge contact configurations to evaluate the effects of contact roughness, pressure and long-time service conditions.

4. Interpretation of test data in terms of generator steady-state capability with unbalanced load.

The second major problem associated with single-phase loads is voltage-unbalance effect on 3-phase motors. The unbalanced voltages, which result from unbalanced currents in the transmission system, when applied across the terminals of a motor, cause double-frequency currents to be induced in the machine rotor. This, as in the case of generators, causes a heating effect which can be injurious. This problem is discussed in further detail in the results.

Proposed Method of Power Supply

To evaluate the current and voltage unbalance problems associated with single-phase railroad electrification, a study was made of that portion of the Norfolk and Western Railway facilities located between Roanoke, Virginia, and Williamson, West Virginia. Profiles of railroad rights of way, together with dispatcher train sheets, for a representative day of heavy traffic movement, were supplied by the Norfolk and Western Railway. An analysis of these data by Gibbs & Hill indicated a peak demand of approximately 80,000 kw, and a load factor of 68%. To serve this load, Gibbs & Hill recommended that the catenary system (50-kv phase-to-phase and 25-kv phase-to-rail) be supplied at six locations, namely: Salem (Roanoke), Walton (Claytor), Glen Lyn, Switchback, Iaeger, and Williamson (Sprigg); see Fig. 2. These power supply points were chosen to be advantageous to both the Norfolk and Western Railway and the American Gas and Electric transmission system.

The American Gas and Electric system assumed in this study was that anticipated for 1958, which included the Clinch River 450,000-kw steam plant. In addition, the effect of all 132-kv interconnections was included.

To serve each of the six Norfolk and Western supply points, step-down transformation was provided directly from the 132-kv system with the exception of Iaeger, where the load was served from the subtransmission network. The latter was chosen to obtain a comparison between the resulting negative-sequence voltages when the single-phase load is served from either the transmission (132-kv) or subtransmission (66-kv) systems.

Since the railroad load was assumed to be served at several locations, a degree of balancing can be obtained by isolating the sections of catenary between adjacent supply points and serving them from different phases. Thus, by using standard transformers, and by designating the catenary system as east and west at each supply point, one possible combination of phasing for the step-down transformers would be as shown in Table I. Many additional combinations of phasing exist but, at this time, only the one combination was studied.

In addition to investigating standard transformation, Scott-connected transformers were studied. Because of the necessity for similar phasing on sections of catenary between adjacent supply points, only two combinations of Scott transformation are possible. Also, since both combinations will give identical results, only the combination shown in Table II was studied.

Method of Analysis

Since the problem of supplying a single-phase load from a 3-phase power system is primarily machine heating (system generators or customer 3-phase motors) due to negative-sequence currents or voltages, and since this heating is a function of time, the simultaneous railroad load at each of the six supply points was calculated every six minutes for a typical 24-hour period. This involved 2,880 individual kilowatt demands at an average power factor of 85%.

In order to relate these known load data with the unknown negative-sequence currents at major generating plants and interconnection points and the negative-sequence voltages at critical station locations (see Fig. 3), equations of the form $I = EY$ were written for the negative-sequence network of the system. The system, as shown in Fig. 3, was reduced to a 23-point equivalent. In these equations, the negative-sequence currents at the six load points were known along with the negative-sequence generator voltages which are zero. To facilitate the solution of the resultant 23 simultaneous equations 240 times (loads at 6-minute intervals over a 24-hour period), the International Business Machine Corporation *650* digital computer was used. Admittance data for the system were obtained from measurements on the A-C Network Analyzer.

Results of Analysis

To relate the resultant unbalanced voltages and currents to heating ($I^2R \times$ Time), the results are presented in the form of cumulative curves in Figs. 4–6. Fig. 4 shows the percentage of time, during a 24-hour period, that the negative-sequence voltages are below the values indicated on the ordinates. With the exception of Iaeger, the negative-sequence voltages at all points are below 1% of rated system voltage for 100% of the time and are below 0.5% for 90% of the time. The unbalance at Iaeger easily could be reduced below 1% if the supply to the railroad load at this point had been at 132 kv instead of at subtrans-

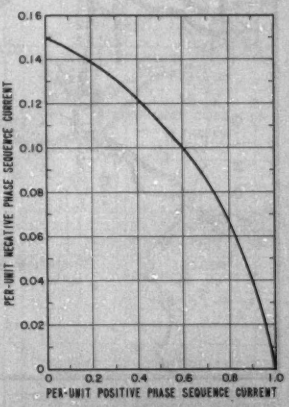


Fig. 1. Permissible negative phase sequence current for turbine-generators[4]

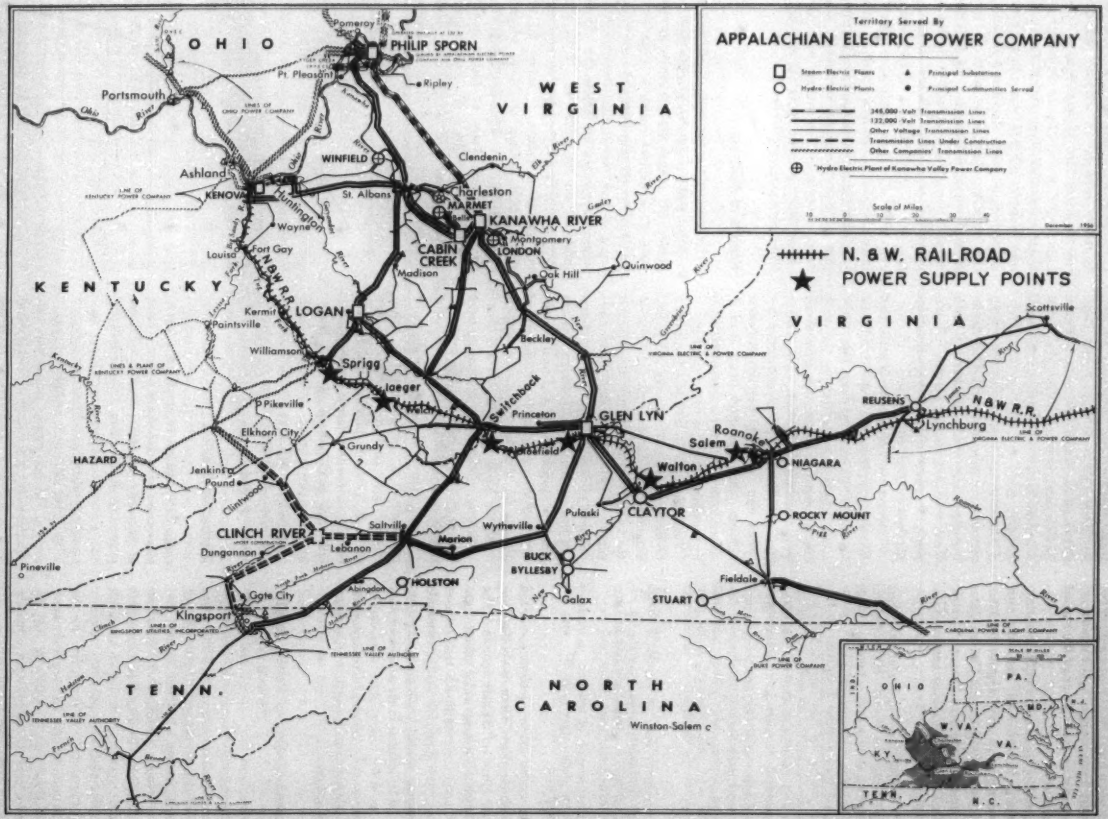


Fig. 2. System map

Table I. Supply System 2-Winding Transformer Phasing

Supply Point	Catenary System	Phasing
Roanoke	East	A-B
	West	C-A
Claytor	East	C-A
	West	B-C
Glen Lyn	East	B-C
	West	A-B
Switchback	East	A-B
	West	C-A
Iaeger	East	C-A
	West	B-C
Sprigg	East	B-C
	West	A-B

Table II. Supply System Scott Transformer Phasing

Supply Point	Catenary System	Phasing
Roanoke	East	Teaser
	West	Main
Claytor	East	Main
	West	Teaser
Glen Lyn	East	Teaser
	West	Main
Switchback	East	Main
	West	Teaser
Iaeger	East	Teaser
	West	Main
Sprigg	East	Main
	West	Teaser

mission voltage. It is believed that the degree of voltage unbalance, shown in Fig. 4, will not cause any difficulty with customer 3-phase motors. Motor heating varies approximately in direct relation to twice the square of the negative-sequence voltage. National Electrical Manufacturers Association standards indicate that a continuous 3.5% negative-sequence voltage will produce a 25% increase in rotor heating. Hence an unbalance of 1% will increase rotor heating by only 2.5% for 10% of the 24-hour period studied. In addition, the negative-sequence voltages, as calculated and plotted in Fig. 4, will be reduced further at customer 3-phase motor terminals due to the shunting-effect of other rotating loads on the system, which effect was not included in the study.

Figs. 5 and 6 indicate the negative-sequence currents as a per-cent of generator ratings at the major power plants. As shown, all negative-sequence currents are reasonable and are below 1% of rated full-load current of the generators for 100% of the time, with the exception of Claytor, Logan, and Glen Lyn. Unbalance at Claytor is not considered excessive, since it is a hydro plant and the machines contain damper windings. The worst unbalance at Logan is on the two 5,000-kw generators while at Glen Lyn it is on the No. 6 high-pressure 125,000-kw unit. At Logan, the negative-sequence current was in excess of 1% of rating for 30% of the time, reaching a peak of 2.5%. At Glen Lyn, the current was in excess of 1% of rating for 65% of the time, reaching a peak of 5 to 6%. The unbalance is more severe at Glen Lyn, since four of the six railroad supply points (Roanoke, Claytor, Glen Lyn, and Switchback) are electrically closer to Glen Lyn than any other major plant. This problem will be reduced considerably by contemplated future transmission and plant additions in this area.

As shown in Figs. 4–6, Scott-connected transformers provide a somewhat better degree of balancing than the arrangement of standard transformers considered in this study at all points, except Glen Lyn.

Conclusions and Recommendations

The foregoing analyses of the unbalanced voltage and current problems associated with serving a single-phase railroad load, like that of the Norfolk and Western Railway, indicate the following:

1. The results show promise for serving single-phase railroad loads by direct transformation without extensive use of phase-balancing equipment.

2. Although the Scott-connected transformer provides a balancing effect at any one supply point, it lacks the flexibility of direct transformation when serving loads at several supply points. A greater measure of over-all balancing may be obtained with direct transformation by the judicious choice of phase combinations for serving successive catenary sections.

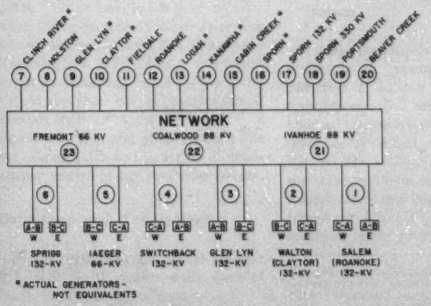


Fig. 3. Equivalent system

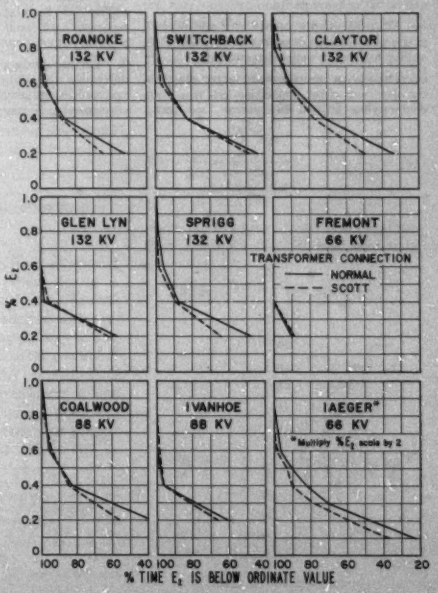


Fig. 4 (right). Negative sequence voltage versus time

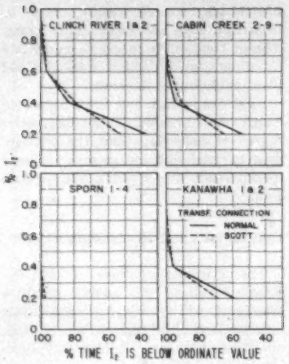


Fig. 5 (left). Negative sequence current versus time

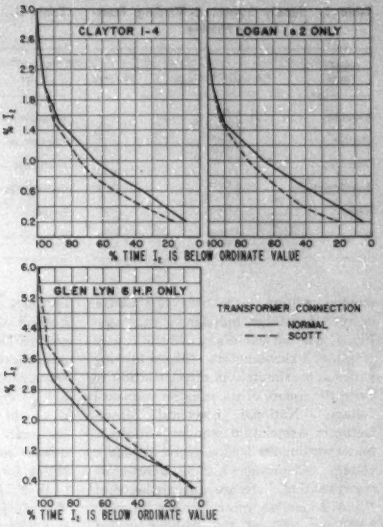


Fig. 6 (right). Negative sequence current versus time

3. The results obtained in this study for the standard transformer connection are based on a particular combination of phasing at the six supply points. Other combinations of phasing, not attempted in this study, might further reduce the extent of unbalance.

4. It appears preferable to serve railroad loads directly from the transmission system, rather than from the subtransmission system. Although this is contrary to some European practice, it has the effect of limiting unbalanced voltages on customer 3-phase equipment and concentrating the problem at generating plants where remedial action can be taken more easily.

5. Multiple supply points and additional single-phase railroad loads will increase the ability to balance generator currents.

6. The present trend toward larger generator sizes and plant capacities, and the accompanying growth in system load and transmission facilities, will greatly enhance the ability to handle single-phase railroad electrification loads.

7. The need to acquire data relative to the ability of turbogenerators to carry unbalanced load remains one of the most important requirements.

8. Further study should be undertaken to determine the shunting effect of rotating system load on generator negative-sequence currents.

References

1. Operation of Large Steam Turbine Generators, J. H. Carter, R. E. Gorman. *AIEE Transactions*, vol. 75, pt. III, Apr. 1956, pp. 217–26.

2. Power Supply for Commercial-Frequency Railroad Electrification, J. C. Price. *Ibid.*, pt. II (Jan. 1957 section), pp. 432–45.

3. A Reappraisal of the Economics of Railway Electrification: How, When and Where Can It Compete with the Diesel-Electric Locomotive? H. F. Brown, R. L. Kimball. *Ibid.*, vol. 73, pt. II, Mar. 1954, pp. 35–51.

The Effect of Quantization in Sampled-Feedback Systems

J. E. BERTRAM
ASSOCIATE MEMBER AIEE

THE RAPID ADVANCE of digital computers and digital technology has resulted in recent years in the development of digital versions of practically every control system component. As a result, mixed systems composed of both digital and analog elements are becoming quite common.

This paper concerns the effect of quantization or round-off errors, inherent in the digital elements, on the performance of feedback control systems. As far as the author has been able to determine there is nothing in the control system literature concerning this problem. However, a rather extensive literature on quantization error is to be found in the fields of pulse code modulation[1,2] and the numerical solution of differential equations.[3-6] In the former, the effort has been directed primarily toward the determination of the mean-squared error resulting from an analog-to-digital conversion in an open-loop communication channel. It has not proven possible in general to extend these results to the feedback case.[7,8] The methods employed in bounding the round-off errors[3-6] in the solution of ordinary differential equations are similar to the methods used in this paper.

Any upper-bound on the over-all effect of a large number of quantizing operations, such as will arise in a feedback system with internal quantization, no matter how precise, is likely to be very conservative because of the more or less random fluctuation in the sign of the error. However, it is believed that the bound serves practically in the comparison of the error arising from different computational schemes such as might be found in a digital controller and gives a reasonable engineering guide to the degree of quantization needed in any particular system.

Definitions and Terminology

Before considering the actual problem it is desirable to define briefly some of the terms which shall be used.

1. *Quantization.* The process of quantizing or converting a signal in analog form to a digital approximation may be illustrated by means of the transfer characteristic of Fig. 1. The range of input magnitudes is divided into disjoint intervals h_i, which are not necessarily equal. All magnitudes falling within each interval are equated to a single value within the interval. This single value is the digital approximation of the magnitude of the analog input signal. Thus if x is the analog input, the digital approximation is given by $y = Q(x)$ where Q is the quantizing function.

2. *A transducer* is considered to be any device which receives an input signal $x(t)$ and produces an output signal $y(t)$ which in general is a function of the past history of its input $x(t)$.

3. *An analog transducer* is a transducer in which the input and output signals are continuous functions of time; the magnitude of these signals may be any value within the physical limitations of the system, i.e., the signals are unquantized.

4. *A sampled-data transducer* is a transducer in which the input and output signals occur only at discrete instants of time (usually periodic) but the magnitudes of the signals as in the analog transducer are unquantized.

5. *A digital transducer* is one in which the input and output signals occur only at discrete instants of time and the signal magnitudes are quantized, i.e., they can assume only certain discrete levels.

On the basis of these definitions a signal which is unquantized in magnitude is often spoken of as an analog signal while if quantized in magnitude it is called a digital signal.

Linear Sampled-Data Feedback System

A linear sampled-data feedback control system is in general composed of a collection of analog and sampled-data transducers. The behavior of such systems for the present problem may be best described by making use of the concept of *state*.[9] The value of the components of the state vector of a transducer at any instant represents that past history of a transducer which is necessary to a determination of its present output and its future state. Using the concept the equations describing the system are:

$$\mathbf{x}(k+1) = A(T)\mathbf{x}(k) + r(k)\mathbf{h}(T) \quad (1)$$

$$c(k) = \mathbf{b}'\mathbf{x}(k) \quad (2)$$

where $\mathbf{x}(k) = \mathbf{x}(kT)$ is an n-dimensional vector with components $(x_1, x_2, \ldots, x_n)$ representing the state of the system at the kth sampling instant, T being the sampling period; $A(T)$ is an $n \times n$ matrix whose elements are functions of the sampling period T; and $\mathbf{h}(T)$ is a vector which represents the effect on the state vector of the input or reference signal $r(t)$. Equation 2 indicates that the output $c(k)$ is a linear combination of the various state variables; the exact combination being given by the column vector $\mathbf{b}$ whose transpose is $\mathbf{b}'$.

When the system equations are written in this manner the characteristic polynomial is the determinant

$$\det[A(T) - I\lambda] = 0 \quad (3)$$

where I is the unit matrix. The necessary and sufficient conditions[10] for the stability of the system are that all n roots, λ_i, of equation 3 have absolute magnitudes less than unity, i.e., $|\lambda_i| \leq 1$ for $i = 1, 2, \ldots, n$.

Since in the present problem it is not necessary that the state variables be physically accessible quantities, the easiest way of obtaining equation (1) is to first obtain the pulse transfer function of all analog transducers in the system. These transfer functions for the linear lumped parameter elements considered will be of the form

$$G(z) = \frac{C(z)}{R(z)} = \frac{\sum_{i=0}^{q} \alpha_i z^{-i}}{1 + \sum_{i=1}^{p} \beta_i z^{-i}} \quad (4)$$

where the β_i's and α_i's are real numbers. Such a transfer function may be simply

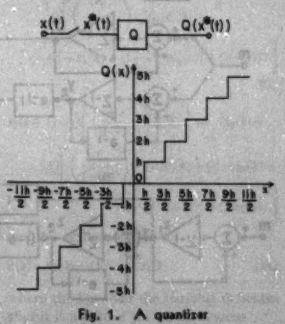


Fig. 1. A quantizer

Paper 58-91, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 15, 1957; made available for printing January 8, 1958.

J. E. BERTRAM is with Columbia University, New York, N. Y.

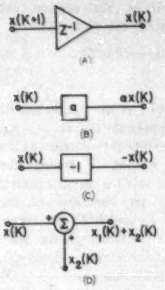


Fig. 2 (left). Basic elements

A—Ideal unit delay
B—Scaler
C—Sign changer
D—Summer

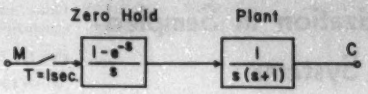


Fig. 3 (above). An example

viewed as a shorthand notation for the linear difference equation with real constant coefficients relating the input $r(i)$ and the output $c(i)$ at sampling instants.

$$c(k)=\sum_{i=0}^{\hat{n}}\alpha_i r(k-i)-\sum_{i=1}^{\hat{q}}\beta_i c(k-i)$$

Once this step is completed, for the purpose of analysis the system may be considered to consist entirely of a collection of sampled-data transducers.

The next step is to select the state variables, i.e., the components of the state vector, and redraw the block diagram of the system in such a way as to make them accessible. This can be done by decomposing the system into the basic elements illustrated in Fig. 2. These elements are the ideal delay element, which delays the input by one sampling period T, a scaler which multiplies the input by a constant, a sign changer which multiplies the input by minus one, and a summer which adds the inputs. There are many ways of performing this decomposition. To illustrate several, consider as an example the plant-hold combination of Fig. 3 whose pulse-transfer function is

$$G(z)=\frac{C(z)}{M(s)}=\frac{e^{-1}z^{-1}+(1-2e^{-1})z^{-2}}{1+(-1-e^{-1})z^{-1}+e^{-1}z^{-2}} \quad (5)$$

1. A canonical form for equation 5 is shown in Fig. 4(A). That this structure has the transfer function of equation 4 is easily checked. This particular structure has the advantage that the values of the various scalar elements are simply the coefficients of z^{-1} in equation 5. The output of each delay element may be considered a component of the state vector so that equations 1 and 2 are obtained by inspection of Fig. 3(A). The state variable x_1 at the $(k+1)$st sampling instant is the input to the first delay element at the kth sampling instant, while state variable x_2 at the $(k+1)$st instant is the x_1 at the kth instant. Thus the state equations are given by

$$x_1(k+1)=(1+e^{-1})x_1(k)+(-e^{-1})x_2(k)+m(k) \quad (6)$$

$$x_2(k+1)=x_1(k)$$

and the output $c(k)$ is

$$c(k)=(e^{-1})x_1(k)+(1-2e^{-1})x_2(k) \quad (7)$$

so that

$$A^{(a)}(T)=\begin{Vmatrix}(1+e^{-1}) & (-e^{-1})\\ 1 & 0\end{Vmatrix} \quad (8)$$

$$\mathbf{h}^{(a)}(T)=\begin{vmatrix}0\\1\end{vmatrix} \text{ and } \mathbf{b}^{(a)}\begin{vmatrix}(e^{-1})\\(1-2e^{-1})\end{vmatrix} \quad (9)$$

2. Another structure is obtained by expanding equation 5 into partial fractions such that

$$G(z)=\frac{z^{-1}}{1-z^{-1}}-\frac{(1-e^{-1})z^{-1}}{1-e^{-1}z^{-1}} \quad (10)$$

Fig. 4(B) shows equation 10 in block diagram form. The state variables are again selected as the outputs of the delay elements so that equations 1 and 2 for this structure are

$$y_1(k+1)=y_1(k)+m(k)$$

$$y_2(k+1)=(e^{-1})y_2(k)+m(k) \quad (11)$$

and

$$c(k)=y_1(k)+(e^{-1}-1)y_2(k) \quad (12)$$

Thus for this structure

$$A^{(b)}(T)=\begin{Vmatrix}1 & 0\\ 0 & e^{-1}\end{Vmatrix} \quad (13)$$

$$\mathbf{h}^{(b)}(T)=\begin{vmatrix}1\\1\end{vmatrix} \text{ and } \mathbf{b}^{(b)}=\begin{vmatrix}1\\(e^{-1}-1)\end{vmatrix} \quad (14)$$

This structure leads to the matrix $A(T)$ being diagonal but has the disadvantage that it requires that the denominator of $G(z)$ be factored.

3. A third possibility is to cascade the basic elements as shown in Fig. 4(C). The state variables are again the outputs of the delay elements so that the system equations are

$$z_1(k+1)=z_1(k)+m(k)$$

$$z_2(k+1)=z_1(k)+e^{-1}z_2(k) \quad (15)$$

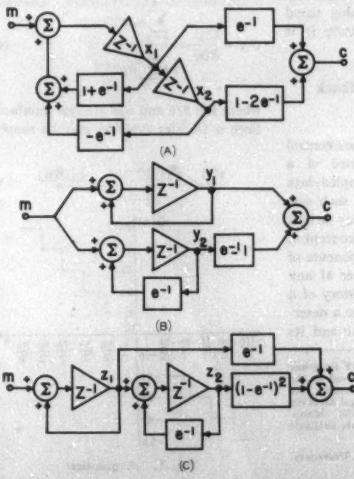


Fig. 4 (left). Decomposition structures

A—Canonical form
B—Partial fraction form
C—Cascaded form

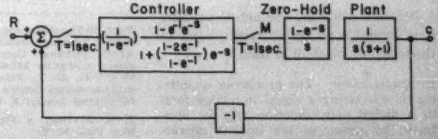


Fig. 5 (below). Feedback example

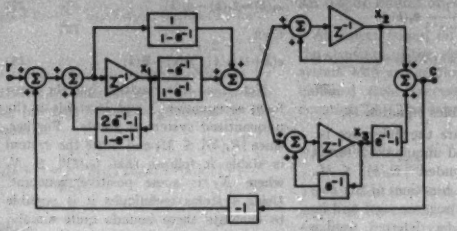


Fig. 6. Decomposition of example

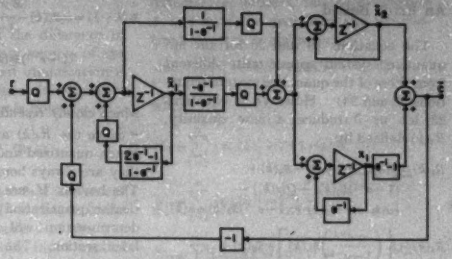


Fig. 7. Example with quantization

and

$$c(k) = \varepsilon^{-1}x_1(k) + (1-\varepsilon^{-1})^2 x_2(k) \tag{16}$$

Therefore, for this structure

$$A^{(c)}(T) = \begin{vmatrix} 1 & 0 \\ 1 & \varepsilon^{-1} \end{vmatrix}$$

$$\mathbf{h}^{(c)}(T) = \begin{vmatrix} 1 \\ 0 \end{vmatrix} \text{ and } \mathbf{b}^{(c)} = \begin{vmatrix} \varepsilon^{-1} \\ (1-\varepsilon^{-1})^2 \end{vmatrix} \tag{17}$$

It is apparent from this example that the decomposition into basic elements is not unique and thus it follows that the selection of state variables is also not unique. However, it should be noted that x, y, and z are related by linear transformations so that the characteristic equations for (a), (b) and (c) are identical so that

$$\det(A^{(a)}(T) - I_\lambda) = \det(A^{(b)}(T) - I_\lambda) = \det(A^{(c)}(T) - I_\lambda) \tag{18}$$

Furthermore the components of the state vector may always be selected as the output of each delay element in the decomposed system, where the number of delay elements is always equal to the $\max(p, q)$ of equation 4.

Once all the elements of a sampled-data feedback system are reduced to this form, the over-all system equations may be written by inspection as was previously done for a single element. As an example consider the system of Fig. 5. It is a typical error-sampled feedback system which is compensated by means of a sampled-data controller in the error line. In this case the controller has been designed to give ripple-free, optimum performance.[9] That is the system has been designed to go to equilibrium with zero system error in $2T$ seconds following a disturbance or step change at the reference input. The plant-hold combination is the same as that of the previous example.

The structure shown in Fig. 6 is the result of using the decomposition of equation 6 for the plant-hold combination and the canonical structure of (a) for the controller. There are three delay elements in this system so that the state vector has three components. The system equations for this example, obtained by inspection are

$$\begin{aligned} x_1(k+1) &= \alpha_1 x_1(k) - x_2(k) + (1-\varepsilon^{-1})x_3(k) + r(k) \\ x_2(k+1) &= \alpha_2 x_1(k) - \frac{\varepsilon^{-1}}{1-\varepsilon^{-1}} x_3(k) + x_3(k) + \frac{1}{1-\varepsilon^{-1}} r(k) \\ x_3(k+1) &= \alpha_3 x_3(k) - \frac{1}{1-\varepsilon^{-1}} x_3(k) + (1+\varepsilon^{-1})x_3(k) + \frac{1}{1-\varepsilon^{-1}} r(k) \end{aligned} \tag{19}$$

where

$$\alpha_1 = \frac{2\varepsilon^{-1}-1}{1-\varepsilon^{-1}}, \quad \alpha_2 = \frac{\varepsilon^{-1}+\varepsilon^{-2}-1}{(1-\varepsilon^{-1})^2}$$

the output is

$$c(k) = x_2(k) + (\varepsilon^{-1}-1)x_3(k) \tag{20}$$

so that

$$A(T) = \begin{vmatrix} \alpha_1 & -1 & 1-\varepsilon^{-1} \\ \alpha_2 & -\frac{\varepsilon^{-1}}{1-\varepsilon^{-1}} & 1 \\ \alpha_3 & -\frac{1}{1-\varepsilon^{-1}} & 1+\varepsilon^{-1} \end{vmatrix} \tag{21}$$

$$\mathbf{h}(T) = \begin{vmatrix} 1 \\ \frac{1}{1-\varepsilon^{-1}} \\ \frac{1}{1-\varepsilon^{-1}} \end{vmatrix} \text{ and } \mathbf{b} = \begin{vmatrix} 0 \\ 1 \\ \varepsilon^{-1}-1 \end{vmatrix} \tag{22}$$

Mixed Feedback Systems

Now if in the example some of the elements are replaced with their digital counterparts the system is no longer linear since the operation of quantization or rounding-off which is inherent in the digital transducer is fundamentally nonlinear.

The digital transducer which replaces a linear sampled-data transducer must perform the same basic operations, that of unit delay or storage, addition, sign change and multiplication by a constant. Since the signal at all points in the digital transducer is in digital and not analog form it is necessary in the mathematical model of the digital transducer to introduce quantizers similar to those shown in Fig. 1 at appropriate points in the system. The correct model, of course, depends on the logical organization of the particular digital transducer being considered.

Now, if in the running example it is assumed that the error sensor and controller are replaced with their digital counterparts whose operations are organized such that the quantizers are located as shown in Fig. 7, one may again describe the system by state equations, which while they are not linear closely resemble the linear system equations 1 and 2.

Again considering the state variables to be the outputs of the delay elements the state equations are

$$\begin{aligned} \bar{x}_1(k+1) &= Q(\alpha_1 \bar{x}_1(k) + Q[-\bar{x}_2(k) + (1-\varepsilon^{-1})\bar{x}_3(k)] + Q[r(k)] \\ \bar{x}_2(k+1) &= Q\left[\left(\frac{-\varepsilon^{-1}}{1-\varepsilon^{-1}}\right)\bar{x}_3(k)\right] + \bar{x}_3(k) + Q\left[\left(\frac{1}{1-\varepsilon^{-1}}\right)(Q(\alpha_2 \bar{x}_1(k) + Q[-\bar{x}_2(k) + (1-\varepsilon^{-1})\bar{x}_3(k)] + Q(r(k)))\right] \\ \bar{x}_3(k+1) &= Q\left[\left(\frac{-\varepsilon^{-1}}{1-\varepsilon^{-1}}\right)\bar{x}_3(k)\right] + \varepsilon^{-1}\bar{x}_3(k) + Q\left[\left(\frac{1}{1-\varepsilon^{-1}}\right)(Q(\alpha_3 \bar{x}_1(k) + Q[-\bar{x}_2(k) + (1-\varepsilon^{-1})\bar{x}_3(k)] + Q(r(k)))\right] \end{aligned} \tag{23}$$

and the system output is

$$\bar{c}(k) = \bar{x}_2(k) - (1-\varepsilon^{-1})\bar{x}_3(k) \tag{24}$$

where the bar over the variable indicates that it is a part of the mixed system.

An Error Bound

The equations 19 and 20 for the unquantized system appear quite different from those of the quantized system (equations 23 and 24). However, in equations 23 let us introduce a new quantity, $R_i(k)$ defined by

$$R_1(k) \triangleq Q(\alpha_1 \bar{x}_1(k)) + Q[-\bar{x}_2(k) + (1-e^{-1})\bar{x}_2(k)] + Q[r(k)] - [\alpha_1 \bar{x}_1(k) - \bar{x}_2(k) + (1-e^{-1})\bar{x}_2(k) + r(k)]$$

$$R_2(k) \triangleq Q\left[\left(\frac{-e^{-1}}{1-e^{-1}}\right)\bar{x}_1(k)\right] + \bar{x}_2(k) + Q\left(\left(\frac{1}{1-e^{-1}}\right)[Q(\alpha_1 \bar{x}_1(k) + Q(-\bar{x}_2(k) + (1-e^{-1})\bar{x}_2(k)) + Q(r(k)))]\right) - \left[(\alpha_2 \bar{x}_1(k)) - \frac{1}{1-e^{-1}}\bar{x}_2(k) + \bar{x}_3(k) + \frac{1}{1-e^{-1}} r(k)\right]$$

and

$$R_3(k) \triangleq Q\left[\left(\frac{-e^{-1}}{1-e^{-1}}\right)\bar{x}_1(k)\right] + e^{-1}\bar{x}_3(k) + Q\left[\left(\frac{1}{1-e^{-1}}\right)(Q(\alpha_3 \bar{x}_1(k)) + Q(-\bar{x}_2(k) + (1-e^{-1})\bar{x}_2(k)) + Q(r(k)))\right] \ldots - \left[\alpha_3 \bar{x}_1(k) - \frac{1}{1-e^{-1}}\bar{x}_2(k) + (1+e^{-1})\bar{x}_3(k) + \frac{1}{1-e^{-1}} r(k)\right]$$

Utilizing this new quantity equation 23 may be written in the form

$$\bar{x}_1(k+1) = \alpha_1 \bar{x}_1(k) - \bar{x}_2(k) + (1-e^{-1})\bar{x}_2(k) + R_1(k) + r(k)$$

$$\bar{x}_2(k+1) = \alpha_2 \bar{x}_1(k) - \frac{e^{-1}}{1-e^{-1}}\bar{x}_2(k) + \bar{x}_3(k) + R_2(k) + \frac{1}{1-e^{-1}} r(k)$$

$$\bar{x}_3(k+1) = \alpha_3 \bar{x}_1(k) - \frac{1}{1-e^{-1}}\bar{x}_2(k) + (1+e^{-1})\bar{x}_3(k) + R_3(k) + \frac{1}{1-e^{-1}} r(k) \quad (25)$$

which closely resembles equation 18.

Since the $R_i(k)$ are the difference between quantized and unquantized terms, they are always bounded, $|R_i(k)| \leq M_i$. The bounds M_i are functions of the particular quantization scheme used and its determination will be deferred until a later section. The important point to be noted here is that the quantized system can be described by the vector equation

$$\bar{\mathbf{x}}(k+1) = A(T)\bar{\mathbf{x}}(k) + \mathbf{R}(k) + r(k)\mathbf{h}(T) \quad (26)$$

where the $R_i(k)$ are simply treated as the components of a disturbance or input vector $\mathbf{R}(k)$. As was previously pointed out the stability of such a system is determined by the matrix $A(T)$ so that if the unquantized system is stable so is the quantized system. Stability, as used in this context, is defined by stating that a system is stable if the response to a bounded input is bounded. In a linear system governed by difference equations the necessary and sufficient condition for this stability as previously mentioned is that the roots λ_i of the determinantal equation $|A(T) - I\lambda| = 0$ must have a magnitude $|\lambda_i| \leq 1$. Since the fictitious input vector to this system is bounded, i.e., $|R_i(k)| \leq M_i$, then if the system is stable all components of the state vector $\bar{\mathbf{x}}$ are bounded and the output $c(k)$ which is a linear combination of these states is also bounded.

Actually, it is of more interest to obtain a bound on the difference between the unquantized system of equation 1 and the quantized system of equation 26. To do this let

$$e_i(k) = \bar{x}_i(k) - x_i(k) \quad (27)$$

then

$$\epsilon(k+1) = A(T)\epsilon(k) + \mathbf{R}(k) \quad (28)$$

Note that this equation has the same form as equation 1 and is stable if the unquantized system is stable. Further, since $|R_i(k)| \leq M_i$ and since the system is stable it follows that $|\epsilon_i(k)| \leq N_i$ where N_i is some positive constant. Using systems techniques it is possible to evaluate these bounds quite simply. From the running example the error equations become

$$\epsilon_1(k+1) = \alpha_1 \epsilon_1(k) - \epsilon_2(k) + (1-e^{-1})\epsilon_2(k) + R_1(k)$$

$$\epsilon_2(k+1) = \alpha_2 \epsilon_1(k) - \frac{e^{-1}}{1-e^{-1}}\epsilon_2(k) + \epsilon_3(k) + R_2(k)$$

$$\epsilon_3(k+1) = \alpha_3 \epsilon_1(k) - \frac{1}{1-e^{-1}}\epsilon_2(k) + (1+e^{-1})\epsilon_3(k) + R_3(k) \quad (29)$$

and the error in the output is

$$\bar{c}(k) - c(k) = \epsilon_2(k) + (e^{-1}-1)\epsilon_3(k) \quad (30)$$

The system described by equations 29 and 30 is shown in block diagram form in Fig. 8 where the $R_i(k)$ are the differences between the state variables of the quantized and unquantized system at the $(k+1)$ sampling instant. Obviously if the R_i were known as a function of k then the differences ϵ_i could be simply calculated. However, the only information which is easy to obtain is the maximum magnitude of the R_i, so that the best that can be obtained is an upper bound on differences E_i.

Before proceeding to the evaluation of this bound let us first summarize our result.

If a linear sampled-data control system, governed by the equations

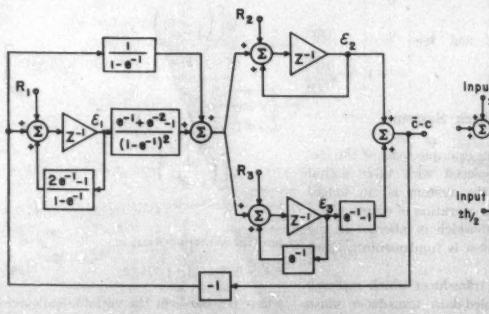


Fig. 8. The difference system

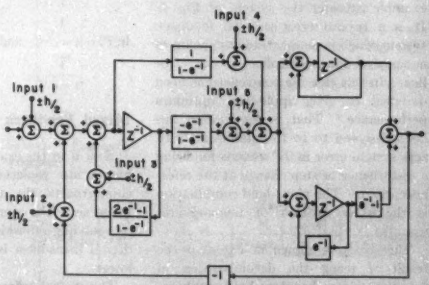


Fig. 9. Upper bounds on $R_i(k)$

$$x_i(k+1)=\sum_{j=1}^{n}\alpha_{ij}(T)x_j(k)+r(k)h_i(T) \text{ for } (i=1, 2, \ldots, n)$$

is stable, then the system obtained by arbitrarily introducing quantizers into the previous system is governed by the equation

$$\mathfrak{x}_i(k+1)=\sum_{j=1}^{n}\alpha_{ij}(T)\mathfrak{x}_j(k)+r(k)h_i(T)+R_i(k) \text{ for } (i=1, 2, \ldots, n)$$

where

$$R_i(k)=Q_i(\mathfrak{x}(k), r(k))-\sum_{j=1}^{n}\alpha_{ij}(T)\mathfrak{x}_j(k)-h_i(T)r(k)$$

which is also stable. Furthermore the difference $\epsilon_i(k)=\mathfrak{x}_i(k)-x_i(k)$ is bounded.

An Upper Bound on $R_i(k)$

In order to bound the difference ϵ_i, it is first necessary to obtain a bound on the components of the disturbance vector $\mathbf{R}(k)$. If we note from the definition of $R_i(k)$, that it is simply the difference between the state vector component, $x_i(k+1)$, obtained in the quantized and unquantized systems for the same state vector and input applied to each at the k sampling instant. Thus a bound on $R_i(k)$ is simply the greatest error due to quantizing in one sampling interval T. If in Fig. 7 each quantizer is replaced by a unity gain plus an input of $\pm h/2$, where h is the quantizing interval, these bounds are easily determined; see Fig. 9. The value $\pm h/2$, is the maximum error in one quantizing operation Q, for any x. This assumes, of course, that all quantizing steps are equal to h. If this is not the case then the inputs should be the greatest vertical deviations from the unit slope line which is the maximum error. Referring to Fig. 9, the bounds on the various R_i are

$$|R_1(k)| \leq |[|h/2|+|h/2|+|h/2|]| \leq 1.5h$$

$$|R_2(k)| \leq |h/2|+|h/2|+\left|\frac{1}{1-e^{-1}}\right| \times |[|h/2|+|h/2|+|h/2|]| \leq 3.4h$$

$$|R_3(k)| \leq |h/2|+|h/2|+\left|\frac{1}{1-e^{-1}}\right| \times |[|h/2|+|h/2|+|h/2|]| \leq 3.4h$$

These results are obtained by noting the values of the inputs to all delay elements resulting from the errors arising in the quantizing elements in one sampling period. It should be noted in the present problem that $|R_2(k)|=|R_3(k)|=h+1.6|R_1(k)|$.

An Upper Bound on $\epsilon(k)$

An upper bound on the difference of the elements of the state vectors of the system with and without quantization is obtained from successive iterations of equation 28. The error is given by

$$\epsilon(k)=A^k\epsilon(0)+\sum_{p=0}^{k-1}A^{k-1-p}\mathbf{R}(k)$$

The physical interpretation of an element $\phi_{ij}(k)$ of the matrix A^k is that it is the response at state variable i at the k'th sampling to a unit pulse input applied at the j'th state at time 0, all other states initially at zero. In explicit notation using these elements, the error is given by

$$\epsilon_i(k)=\sum_{j=1}^{n}\phi_{ij}(k)\epsilon_j(0)+\sum_{p=0}^{k-1}\sum_{j=1}^{n}\phi_{ij}(k-1-p)R_j(p)$$

for $i=1, 2, \ldots, n$. Taking the absolute magnitude of both sides of this equation and using the triangular inequality and the fact that $|R_j(k)|<M_j$ for all k, then

$$|\epsilon_i(k)| \leq \sum_{j=1}^{n}|\phi_{ij}(k)||\epsilon_j(0)|+\sum_{j=1}^{n}M_j\sum_{p=0}^{k-1}|\phi_{ij}(k-1-p)|$$

Using this relation it is possible to determine the bound for any value of k and any initial difference $\epsilon(0)$. However, it is often of more interest to determine the bound on the asymptotic behavior as $k\to\infty$.

That is

$$\lim_{k\to\infty}|\epsilon_i(k)| \leq \lim_{k\to\infty}\sum_{j=1}^{n}|\phi_{ij}(k)||\epsilon_j(0)|+\sum_{j=1}^{n}M_j\lim_{k\to\infty}\sum_{p=0}^{k-1}|\phi_{ij}(k-1-p)|$$

For a stable linear system the various unit pulse responses are absolutely convergent, so it follows that the first term on the right

$$\lim_{k\to\infty}\sum_{j=1}^{n}|\phi_{ij}(k)||\epsilon_j(0)|\to 0$$

while the second term

$$\lim_{k\to\infty}\sum_{p=0}^{k-1}|\phi_{ij}(k-1-p)|\to N_{ij}$$

where N_{ij}, some positive constant, is the sum of the absolute magnitudes of the unit pulse response from i to j.

Using these results on the running example, the asymptotic bound on the difference e_1 between the state variables $\mathfrak{x}_1$ and x_1 of the quantized and unquantized systems is given by

$$\lim_{k\to\infty}|e_1(k)| = M_1N_{11}+M_2N_{12}+M_3N_{13}$$

where

$$N_{11}=\sum_{k=0}^{\infty}|\phi_{11}(k)| = 1/(1-e^{-1})$$

$$N_{12}=\sum_{k=0}^{\infty}|\phi_{12}(k)| = 1$$

$$N_{13}=\sum_{k=0}^{\infty}|\phi_{13}(k)| = 1-e^{-1}$$

and since $M_1=1.5h$ while $M_2=M_3=3.4h$, it follows that

$$\lim_{k\to\infty}|e_1(k)| \leq 7.9h$$

Generally the errors will be considerably less than this bound.

Conclusion

In this paper it has been shown that the introduction of digital transducers into a sampled-data system does not result in instability, but it does produce an error as a result of the quantization. This error can be bounded by simply replacing every quantizer with a unit gain element plus an input which is equal in magnitude to the greatest error possible in the element. The magnitude of the output of the system resulting from all of these inputs is an upper bound on the error.

The upper bound obtained in this way is quite conservative since the signs of the error tend to alternate. However, since the result is so easily obtained, it provides a simple means of comparing various computing schemes in the controller and furthermore, serves to select the value of h.

References

1. Spectra of Quantized Signals, W. R. Bennett. *Bell System Technical Journal*, New York, N. Y., vol. 27, July 1948, pp. 446-72.

2. A Study of Rough Amplitude Quantization by Means of Nyquist Sampling Theory, B. Windrow. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-3, Dec. 1956, pp. 266-76.

3. On the Accumulation of Errors in Numerical Integration, Dirk Brouwer. *Astronomical Journal*, New Haven, Conn., vol. 46, 1937, pp. 149-53.

4. On the Accumulation of Errors in Processes of Integration on High-Speed Calculating Machines, Hans A. Rademacher. *The Annals*, Computation Laboratory, Harvard University, Cambridge, Mass., vol. XVI, 1947, pp. 176-87.

5. The Accuracy of Numerical Solutions of Ordinary Differential Equations, Theodore

E. Sterne. *Mathematical Tables and Other Aids to Computation*, Washington, D. C., vol. VII, 1953, pp. 159–64.

6. Planning and Error Considerations for the Numerical Solution of a System of Differential Equations on a Sequence Calculator. F. J. Murray. *Ibid.*, vol. IV, 1950, pp. 133–44.

7. On the Precision of a Certain Procedure of Numerical Integration. H. D. Huskey, D. R. Hartree. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 42, 1949, pp. 57–62.

8. Round-Off Errors in Numerical Integration on Automatic Machinery. G. Forsythe. *Bulletin*, American Mathematical Society, Providence, R. I., vol. 56, 1950, pp. 61–62.

9. General Synthesis Procedure for Computer Control of Single and Multiloop Linear Systems. R. E. Kalman, J. E. Bertram. "Proceedings of the Computers in Control Systems Conference." *AIEE Special Publication T-101*, May 1958, pp. 130–36.

10. On the Boundness of Solutions of Non-Linear Differential and Difference Equations. R. Bellman. *Transactions*, American Mathematical Society, Providence, R. I., vol. 62, 1947, pp. 357–86.

The Sensitivity of the Poles of Linear, Closed-Loop Systems

R. Y. HUANG
NONMEMBER AIEE

Synopsis: One method of evaluating the constancy of dynamic response of a system is to determine how sensitive the poles of the system are to variations of the system parameters. If the sensitivity of a pole to one particular parameter is known, then for small variations of the parameter, the total change in the pole position may be easily computed. This method is used to show how dynamic error analysis in linear computer setups may be accomplished in a straightforward manner. The definition is also of some value as a general tool in the analysis of linear feedback systems.

IN THE APPLICATION of feedback theory, the problems may be divided, in general, into two classes: the first are those encountered in the stabilization of elements (e.g. amplifiers, servos) to be used as parts of a computing loop. The primary objectives of feedback are, to obtain greater static accuracies, and to increase the natural frequencies of these elements, so that they are very large compared to the frequencies in the computation. Usually, the only stipulation concerning damping is that it be "reasonable"; in some cases, an overshoot is specified, merely in the interest of reducing phase error. It must not be thought that such an element must be an individual component. In some cases, it might be what is usually considered a whole computation, containing as parts of itself many amplifiers and servos, as for instance, an automatic process in a factory, because all that may be required is that it be stable, and that its natural frequency be sufficiently larger than the over-all natural computing frequencies of the factory.

The second class of problems is those involved in the application of these elements in a computing loop. Here it is required that the outputs of the system be specified functions of the inputs. If the system is linear, these input-output relations may be expressed in terms of an input-output transfer function. Hence, how closely the system resembles the desired one depends on how closely these input-output transfer functions are reproduced.

This paper is concerned with the problem of predicting the accuracy of reproduction of the dynamics of a linear system, having been given elements whose static and dynamic characteristics are known to given degrees of accuracies; or of elements that are assumed ideal, but are actually not, and of computing the effects of the unidealness of these elements on the dynamics of the system.

For this purpose, what is called the sensitivity of a pole of the closed system is defined. This is the percentage variation of this pole for a given percentage variation of one of the parameters of the system. If the parameter variation is sufficiently small, the total change in the pole position due to this parameter variation may easily be found. This idea will be applied to a simple single-loop system, then to more complex systems, to show how dynamic error analysis in a linear computer setup may be made a matter of routine. At the same time, it is seen that this furnishes a new aspect of some of the problems associated with the stabilization of simple systems. It also provides a particularly simple means of causing a single-loop system to exhibit relatively constant damping in spite of variation of one of its significant time constants.

Paper 58-80, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted June 17, 1957; made available for printing November 29, 1957.

R. Y. Huang is with the Curtiss-Wright Corporation, Carlstadt, N. J.

Sensitivity of a Pole of a Single-Loop System

Given the system which is shown in Fig. 1, let

$$H(s)=\frac{G(s)}{1+G(s)} \tag{1}$$

where

$$G(s)=\frac{K(s+z_1)(s+z_2)\dots(s+z_m)}{(s+\mathcal{P}_1)(s+\mathcal{P}_2)\dots(s+\mathcal{P}_n)} \tag{2}$$

$$H(s)=\frac{K(s+z_1)(s+z_2)\dots(s+z_m)}{(s+p_1)(s+p_2)\dots(s+p_n)} \tag{3}$$

The sensitivity of a particular closed-loop pole p_k to any open-loop parameter k may be defined as

$$S_k^{p_k}=\frac{\dfrac{\partial p_k}{p_k}}{\dfrac{\partial k}{k}}=\frac{k}{p_k}\frac{\partial p_k}{\partial k} \tag{4}$$

and gives the percentage change in p_k for a given percentage change in k. Given equations 1, 2, and 3,

$$(s+\mathcal{P}_1)(s+\mathcal{P}_2)\dots(s+\mathcal{P}_n)+K(s+z_1)(s+z_2)\dots(s+z_m)=(s+p_1)(s+p_2)\dots(s+p_n) \tag{5}$$

For the sake of simplicity of notation, examine $S_{\mathcal{P}_1}^{p_1}$. It is obvious that this applies to any p_k with respect to any $\mathcal{P}_j$. Let $\mathcal{P}_1$ vary so that it becomes $\mathcal{P}_1+\Delta\mathcal{P}_1$. At the same time, to maintain the same amount of feedback, K must vary by the same percentage. K becomes $K(1+\Delta\mathcal{P}_1/\mathcal{P}_1)$. Equation 5 becomes

$$(s+\mathcal{P}_1+\Delta\mathcal{P}_1)(s+\mathcal{P}_2)\dots(s+\mathcal{P}_n)+K\left(1+\frac{\Delta\mathcal{P}_1}{\mathcal{P}_1}\right)(s+z_1)(s+z_2)\dots(s+z_m)=(s+p_1+\Delta p_1)(s+p_2')\dots(s+p_n') \tag{6}$$

where $p_2'\neq p_2;\ \dots p_n'\neq p_n$. Substracting equation 5 from equation 6 and letting

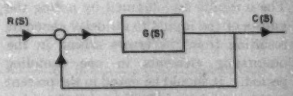


Fig. 1.

$s=-p_1$

$$\Delta\mathcal{P}_1(\mathcal{P}_2-p_1)\ldots(\mathcal{P}_n-p_1)+K\frac{\Delta\mathcal{P}_1}{\mathcal{P}_1}(z_1-p_1)(z_2-p_1)\ldots(z_m-p_1)=\Delta p_1(p_2'-p_1)\ldots(p_n'-p_1)$$

or

$$\frac{\Delta p_1}{\Delta\mathcal{P}_1}=\frac{(\mathcal{P}_2-p_1)\ldots(\mathcal{P}_n-p_1)+\dfrac{K}{\mathcal{P}_1}(z_1-p_1)\ldots(z_m-p_1)}{(p_2'-p_1)\ldots(p_n'-p_1)} \tag{7}$$

as $\Delta\mathcal{P}_1\to 0$, $(p_2'-p_1)\to(\mathcal{P}_2-p_1)$; $(p_n'-p_1)\to({}_np-p_1)$. Also, from equation 5, at $s=-p_1$

$$(\mathcal{P}_2-p_1)\ldots(\mathcal{P}_n-p_1)=-\frac{K(z_1-p_1)(z_2-p_1)\ldots(z_m-p_1)}{\mathcal{P}_1-p_1}$$

equation 7 becomes

$$\frac{\Delta p_1}{\Delta\mathcal{P}_1}=\left(-\frac{1}{\mathcal{P}_1-p_1}+\frac{1}{\mathcal{P}_1}\right)\frac{K(z_1-p_1)(z_2-p_1)\ldots(z_m-p_1)}{(p_2-p_1)\ldots(p_n-p_1)}=-\frac{p_1}{\mathcal{P}_1(\mathcal{P}_1-p_1)}k_{-p_1}$$

where k_{-p_1} denotes the residue of $H(s)$ at $s=-p_1$. Hence

$$S_{\mathcal{P}_1}^{p_1}=\frac{\Delta p_1}{\Delta\mathcal{P}_1}\frac{\mathcal{P}_1}{p_1}=\frac{1}{p_1-\mathcal{P}_1}k_{-p_1} \tag{8}$$

In general, dropping the subscripts

$$S_{\mathcal{P}}^{p}=\frac{1}{p-\mathcal{P}}k_{-p} \tag{9}$$

Similarly, the sensitivities to a zero z and to the loop gain K may be evaluated. Hence

$$S_z^p=-\frac{1}{p-z}k_{-p}$$

$$S_K^p=\frac{1}{p}k_{-p} \tag{10}$$

Over-all Sensitivity to Poles and Zeros

If the following is defined

$$S_{\mathcal{P}_2}^{p}=S_{\mathcal{P}_1}^{p}+S_{\mathcal{P}_2}^{p}+\ldots+S_{\mathcal{P}_n}^{p}+S_{z_1}^{p}+S_{z_2}^{p}+\ldots+S_{z_m}^{p}$$

it may be shown that this is

$$S_{\mathcal{P}_2}^{p}=1 \tag{11}$$

Considered even more generally, it can be seen that this is true. Consider the computer solution of any system. Increasing all the poles and zeros by the same percentage is equivalent to changing the time scale. (Although any singularities at the origin do not actually move, in regard to percentages they must be presumed to have done so, and K changed accordingly.) Hence the over-all sensitivity of any closed-loop pole to all the system poles and zeros is unity.

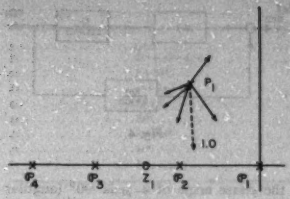


Fig. 2.

An interesting method of graphically determining the sensitivity of a closed-loop pole may be deduced from equation 11. The inverse of the distance from the closed-loop pole to each of the open-loop zeros is reversed in phase, and summed vectorially with the inverses of the distance to the poles. The vector sum is set equal to unity, and the sensitivity of the given closed-loop pole to any open-loop pole or zero may be found by comparing it with this known vector. This method is illustrated as follows: In the pole-zero configuration of Fig. 2, p_1 is one of the closed-loop poles, and the other singularities are open-loop poles and zeros. The vector sum, following the procedure which is described in the above, is given by the dotted line; this is unity. The sensitivity of p_1 to, say, $\mathcal{P}_2$, is the ratio of the vector, denoting the inverse of the distance from p_1, to the vector denoting unity. In this case, $S_{\mathcal{P}_2}^{p_1}\approx 0.7$. In a similar manner, the other sensitivities may be found. This method gives not only the magnitude of the sensitivity, but also the phase.

Variation of Sensitivity

Unlike static sensitivity, where variations are not easily predictable, the variation of the sensitivity of a closed-loop pole as the pole travels along its locus is readily apparent. Starting from its open-loop position with a small amount of feedback, the closed-loop pole is still very close to its original open-loop pole, and will be quite sensitive to it, and only very little to the other singularities. On the other hand, as it approaches another closed-loop pole, the sensitivity of this pole to all the open-loop paramaters becomes very great; if these two poles coincide to form a second-order pole, the sensitivity becomes infinite.

In this sense, critical damping may be considered a highly unstable state, in that a slight variation in one of the open-loop parameters can remove the poles well away from the point of critical damping. This is, however, largely an academic question.

As feedback is increased and the closed-loop pole travels towards its final position, either finite zero or infinity, the sensitivity in the case of the former is great only towards the zero it is approaching. In the case of the latter the sensitivities to the open-loop singularities will be approximately the same if the given pole has a small damping ratio; if it has a large damping ratio, it will tend to be more sensitive to the singularities that are far removed from the origin.

In general, the sensitivity will be constant if the pole is far removed from any region where the root locus has a large curvature, and will be rather variable in the neighborhood of such regions. It is interesting to note that the closed-loop pole is equally sensitive to an open-loop pole as to a zero, but that the motion of the closed-loop pole due to a motion of the open-loop pole, and to a similar motion of the zero, are directly opposite to each other. The sensitivity to a small time constant is approximately proportional to the distance between the time constant and the given closed-loop pole. From this may be made an approximate estimate of the effect of extraneous lags on the closed-loop pole position. Since

$$\frac{\mathcal{P}}{p}\frac{dp}{d\mathcal{P}}=S_{\mathcal{P}}^{p}=\frac{k_{-p}}{p-\mathcal{P}}$$

$$\frac{k_{-p}}{\mathcal{P}(p-\mathcal{P})}d\mathcal{P}=\frac{dp}{p}$$

Both sides integrated, and the assumption made that k_{-p} is a constant for the integration, and also that p is a constant for the integrand on the left, then

$$\frac{k_{-p}}{p}\ln\frac{p-\mathcal{P}}{\mathcal{P}}\Bigg|_{\mathcal{P}_2}^{\mathcal{P}_1}=\ln p\Bigg|_{p_2}^{p_1}$$

$$\left(\frac{p-\mathcal{P}_1}{p-\mathcal{P}_2}\frac{\mathcal{P}_2}{p_1}\right)^{-\frac{k_{-p}}{p}}=\frac{p_1}{p_2}$$

The case is interesting where $\mathcal{P}_1\to\infty$, and hence

$$\left(\frac{\mathcal{P}_2}{\mathcal{P}_2-p_1}\right)^{-\frac{k_{-p_1}}{p_1}}=\frac{p_1}{p_2} \tag{12}$$

or,

$$p_2=p_1\left(\frac{\mathcal{P}_2}{\mathcal{P}_2-p_1}\right)^{\frac{k_{-p_1}}{p_1}}$$

A simpler, but essentially nearly as accurate, approximation may be arrived

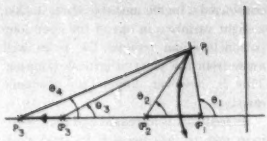


Fig. 3.

at if it is assumed that, for $\mathcal{P}_2$ sufficiently large

$$S_{\mathcal{P}_2}^{p} = -\frac{k_{-p}}{\mathcal{P}_2}$$

If this assumption is made, the following approximation is arrived at

$$p_2 = p_1 \exp \frac{k_{-p1}}{\mathcal{P}_2} \tag{13}$$

The sensitivity to the loop gain K, it will be noted, is identical to the sensitivity to a pole at the origin. The philosophy behind the method of stabilization by pole compensation may be clearly seen from this; for if the pole is sufficiently close to the origin, the effects on the closed-loop poles of varying the pole position and of varying the loop gain are identical, hence one may be "traded" for the other.

An interesting case is the one where the phase angle of the sensitivity of a control pole to one of the open-loop parameters is zero. It is seen that, if the open-loop parameter were a pole or a zero, and it changed position on the real axis, or if the open-loop parameter were the loop gain K, and it were increased or decreased, the phase angle of zero would correspond to the constant-damping case as these parameters were varied. The question of whether the effect of changing a particular parameter is "stabilizing" or "destabilizing," at least in the small, depends on whether the phase angle of the sensitivity of the control pole to this parameter is positive or negative as this parameter is varied.

The Constant-Damping Problem

The constant-damping problem is whether or not a closed loop may maintain reasonably constant damping in spite of a given variation of one of its open-loop poles. A solution may be found by determining when the sensitivity of the control pole to this one particular parameter has a zero-phase angle. As an example, the configuration of Fig. 3 is taken, and the location of p_1 at which it has constant damping with respect to small variations in $\mathcal{P}_2$ is to be found.

Note that the phase angle of $S_{\mathcal{P}_2}^{p_1}$ is θ_2, plus the phase angle of k_{-p1}. But

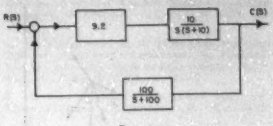


Fig. 4.

the phase angle of k_{-p1} is 90° (angular degrees) $-\theta_4$, hence $\theta_2+\theta_4=90°$. But $\theta_1+\theta_3+\theta_2=180°$, and also $\theta_4=\theta_3$, although slightly smaller than θ_3. Hence θ_1 must be approximately 90°, and will be smaller than 90° by the amount by which θ_2 exceeds θ_4. It may be readily seen that this same solution applies substantially to a system of any higher order, provided that the higher-order singularities are far enough removed from the control poles. It may also be seen that the effect of a zero on the required angle θ_1 is nil, hence a zero acts only to change the possible position of the pole p_1.

To summarize: if there are two dominant poles, and any number of zeros, and if all other poles are several times further out on the real axis, then, in order that constant damping for the control poles when the loop is closed be maintained in spite of variation of one of the two dominant poles, the control pole should be located at the same distance from the imaginary axis as is the other dominant pole. If higher-order poles are closer to the dominant poles, then the control pole should be moved closer to the imaginary axis than should the other dominant pole.

Some Examples of the Application of Sensitivity to a Single-Loop System

Example 1

See the servo system of Fig. 4. It is known that the servo time constant may vary over a large number of units by as much as 10%, while, over a period of time, the loop gain may vary by as much as 6%. What is the total maximum variation in the damping of the control poles that may be expected? What are the maximum and minimum limits of the overshoot in the step-function response?

The closed-loop function is

$$\frac{C}{R}(s) = \frac{92(s+100)}{(s+101)(s+4.5+j8.4)(s+4.5-j8.4)}$$

hence

$$k_{-p1} = j5.47$$

The percentage variation in the control pole due to variation in the motor time constant is 0.0545, and that due to variation in the loop gain is 0.0344. The

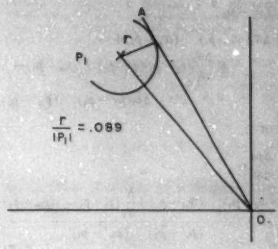


Fig. 5.

total percentage variation is 0.089. As seen in Fig. 5, the concern is to find the angle θ, where the line OA is tangent to the circle. It is obvious that θ is the arctangent of the percentage variation, hence the maximum and minimum damping angles are 67° and 56.8°, respectively. Hence

$$\zeta_{min} = 0.39$$

$$\zeta_{max} = 0.55$$

and the maximum overshoot to a step function would be 27%, while the minimum overshoot would be 11%. (These are really absolute limits, since both the loop gain and the servo time-constant variations must be real.)

Example 2

The following transfer function is to be synthesized:

$$H(s) = \frac{K}{(s+1+j20)(s+1-j20)(s+60+j30)(s+60-j30)}$$

with a single-loop system, using elements of the form $K_1/S+a$. If the individual time constants of each element vary by as much as 1%, the question is whether there is a limit to the amount of feedback that may be allowed while it is assured that, in no case, the system go into oscillation.

The residue at a control pole k_{-p1} is

$$|k_{-p1}| = \frac{K}{1.8\times10^6}$$

If this is to be a single-loop system, and the open-loop poles are to be restricted to the left-half plane, K may not exceed 1.8×10^6. Hence the maximum value of k_{-p1} is

$$|k_{-p1}|_{max} = 10$$

If the open-loop poles are further restricted to be real, then the minimum distance of the control pole to any open-loop pole must be 20; hence the maximum

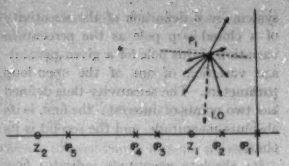


Fig. 6.

Table I.

s_1	$Q(s)$	$Q(-p_1)$	$Q(-p_1)\times$ % error
z_1	$2s^4+s^3+4s+2$	$-6.31+j8.7$	0.74 /140.7°
p_1	$2+5s^3$	$-8.15+j5.05$	0.596 /122°
z_2	s^2+3s+2	$2.33+j4.34$	0.492 /61.9°
s_4	$7s+2$	$5.16+j7.77$	0.934 /56.4°
p_2	$2s^3$	$-3.16-j1.35$	0.685 /−156.9°
p_3	$2s^3+s^2$	$-4.19-j0.84$	0.84 /−125.4°
p_4	$4s^3$	$-4.12+j4.04$	1.15 /135.5°
p_5	$3s$	$1.35+j3.33$	0.719 /68°

sensitivity of the control pole to any of the open-loop poles must not exceed 1/2. If each open-loop pole varies by 1%, each variation may cause a maximum variation of 0.5% of the control pole; thus the effect of all four poles may not exceed 2%, and probably will be much less. It is easily seen that this is not sufficient to cause the control pole to move to the right-half plane. On the other hand, if each open-loop pole may vary by as much as 4%, it might be advisable to restrict the amount of feedback to less than the maximum.

Example 3

The open-loop pole-zero configuration shown in Fig. 6, and one of its closed-loop control poles are given. Indicate which of the open-loop singularities are stabilizing, in the sense that if the singularity is moved towards the origin, the damping of the control poles will increase, and which are destabilizing. Further, if the damping of the control pole is to be increased, indicate which singularity should be moved for a maximum effect.

As may be seen from Fig. 6, the vector summation of the inverse of the distances is labeled unity. All vectors that lead the unity vector by less than 180° are "stabilizing," while those that lag by less than 180° are "destabilizing." Hence $\mathcal{P}_1$, $\mathcal{P}_2$, z_1, and z_2 are stabilizing, while $\mathcal{P}_3$, $\mathcal{P}_4$, and $\mathcal{P}_5$ are destabilizing. If a line is drawn (dotted line in Fig. 6), perpendicular to the unity vector, and the other vectors are projected on this line, the vector with the longest projection should have maximum control on the damping of p_1. In this case, it looks as if $\mathcal{P}_1$, $\mathcal{P}_3$, and $\mathcal{P}_4$ will do equally well in this respect. However, $\mathcal{P}_1$ is at the origin, hence changing it corresponds to decreasing the loop gain K. This is not a satisfactory solution, but $\mathcal{P}_3$ and $\mathcal{P}_4$ may be varied for equal effect, and in this case they should be moved away from the origin.

Example 4

Take the closed-loop function of example 2, where $K=1.8\times10^6$. If there is an extraneous lag in the open loop, how large must it be before the system will go into oscillation?

$k_{-p_1}=-4.95+j8.35$

Using the approximation of equation 13, the location of p_1 will be

$$p_1=(1+j20)\exp\frac{-4.95+j8.35}{\mathcal{P}}$$

where $\mathcal{P}$ is an extraneous lag in the open-loop function. For finding out when p_1 becomes wholly imaginary, only the imaginary component in the exponent of e is of interest. Without the lag, as p_1 is at an angle of 87.14°, the imaginary part of e must contribute an additional 2.86°. Hence

$$\frac{8.35}{\mathcal{P}}=0.05 \text{ radians}$$

and the required lag is

$$\mathcal{P}=167$$

Sensitivity in Multiloop Systems

The method of finding the sensitivity of the closed-loop pole in a single-loop system may be extended as well to finding the sensitivity of a closed-loop pole to any parameter in a multiloop system. If the closed-loop function is written in the form

$$H(s)=\frac{K_1(s^m+a_1s^{m-1}+\ldots+a_m)}{s_n+b_1s^{n-1}+\ldots+b_n+K_s(s+\mathcal{P})(s^q+c_1s^{q-1}+\ldots+c_q)} \quad (14)$$

which is always capable of being done, and if

$$Z(s)=K_1(s^m+a_1s^{m-1}+\ldots+a_m)$$

$$Q(s)=K_s(s+\mathcal{P})(s^q+c_1s^{q-1}+\ldots+c_q) \quad (15)$$

then it may be seen, by a process analogous to those in equations 1 through 8, that

$$S_{\mathcal{P}}^{p}=\frac{1}{\mathcal{P}-p}\frac{Q(-p)}{Z(-p)}k_{-p} \quad (16)$$

Also, let $\mathcal{P}=0$ in equation 14; then equation 16 will give the variation of p_1 with k_s. Hence

$$S_{k_s}^{p}=\frac{1}{p}\frac{Q(-p)}{Z(-p)}k_{-p} \quad (17)$$

It is seen that equations 9 and 10 are subcases of equation 16, and may be derived from it. A simple example is the system illustrated in Fig. 7. It may be shown that, if $\mathcal{P}$ is a pole of $G_1(s)$,

$$S_{\mathcal{P}}^{p}=\frac{1}{p-\mathcal{P}}\frac{1}{1+G_1(-p)}k_{-p} \quad (18)$$

by applying equation 16. In general, equations of the form of equation 18 may be derived from equation 16 for most given configurations, by reapplying the equation $1+G(-p)=0$, where $G(s)$ is closed around itself, and p is a pole of $G(s)/1+G(s)$.

Example 5

See the computer setup shown in Fig. 8. While the integrators and summers may be taken as ideal dynamically, the input resistors are not prefect. Hence it is estimated that g_1, g_3, g_5, and g_4 may be

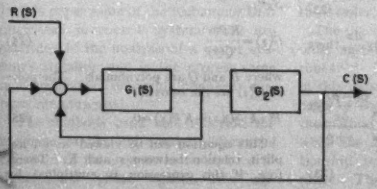


Fig. 7 (left).

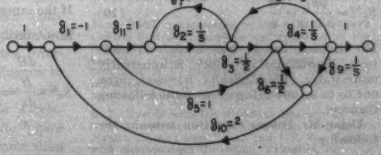


Fig. 8 (right).

off by as much as 0.1%, while g_5, g_6, g_7, and g_8 may be off by as much as 0.2%. What is the maximum per cent of error of the control poles?

The input-output transfer function may be written down by inspection, and, substituting in numerical values

$$\frac{C}{R}(s)=\frac{g_1g_2g_3g_4g_{11}+g_1g_4g_5g_{11}}{1-g_1g_2g_3g_{10}-g_1g_2g_3g_4g_5g_6g_{11}-g_1g_2g_7g_{11}-g_4g_5g_{11}-g_1g_4g_5g_6g_{10}}$$

$$=-\frac{1}{4}\frac{s(1+2s)}{s^2+1.25s^2+1.75s+0.5}$$

One control pole is located at $p_1=0.451+j1.11$. Hence

$$-\frac{1}{p_1}\frac{k-p_1}{z(-p_1)}=-0.298/-4.7°$$

$Q(-p_1)$ must now be found for each individual gain $g_1g_2g_3g_4g_5g_6g_7$ and g_8. Note that $Q(s)$ for any gain g consists of those terms of the denominator which contain g as a product. Hence the $Q(s)$ for each g may be written by inspection, and, substituting in $s=-p_1$, the $Q(-p_1)$ for each g. This is given in Table I.

The problem of obtaining the maximum error corresponds to that of deciding, for each of the vectors in the right-hand column of the table, whether to take the vector as is, or to reverse the vector 180°. It is suggested that a relaxation procedure be resorted to. Initially, each vector may be reversed or not as desired. The summation of these vectors may be called the first sum. Compare each vector to this first sum, and reverse only those vectors that are more than 90° out of phase with the first sum. The summation of the result is the second sum. The procedure is repeated until no vector is 90° out of phase with the nth sum. The nth sum then gives the desired maximum.

With this procedure is obtained for the nth sum

$$\sum_n=-2.22+j3.22=3.91/-145.4°$$

The maximum error is the nth sum, multiplied by $-1/p_1\ \ k-p_1/Z(-p_1)$. Hence

$$\max\text{ error}=(3.91\times0.298)\%=1.16\%$$

The effect on the pole position of an extraneous lag in a multiloop system may be found analogously to equation 13. Here

$$p_2=p_1\exp-\frac{Q(-p_1)}{z(-p_1)}\frac{k-p_1}{\theta_2} \tag{19}$$

With equations 17 and 19, dynamic error analysis in a linear computer setup becomes a straightforward matter. The effect of limited dynamic response of computer components may be simulated by either one or two "extraneous" lags, and a solution may be obtained for any multiloop system similar to that given by example 5.

The effect on zero position of system parameter variation may also be found, using a procedure similar to that for obtaining equation 16.

Conclusions

An attempt has been made to evaluate the constancy of response of a closed-loop system by a definition of the sensitivity of a closed-loop pole as the percentage variation of this pole for a given percentage variation of one of the open-loop parameters. The sensitivity thus defined has two points of interest: the first is its absolute magnitude, and the second is its phase.

Its absolute magnitude indicates how sensitive the pole is to this particular parameter, while the phase indicates the direction of movement of the pole for a given change in the parameter. In analogue computer setups, where the dynamics of a simulated system must be faithfully reproduced, the magnitude term gives an idea of the total possible error in the pole position. It is true, however, that probably certain errors are more objectionable than others. For instance, if the error is in the direction of changing only the natural frequency, this is probably a lot more tolerable than if it were in the direction of changing damping. The phase term indicates exactly in which direction this change occurs, since in practice all variations of open-loop parameters (loop gain or pole or zero position) are restricted to being real.

It would seem that one of the applications of this definition of dynamic sensitivity is in the error analysis of computer setups. The dynamic errors due to incorrect static gain, and to the limited frequency responses of actual computing elements, may be evaluated routinely. The definition also has some intuitive value as a general analytic tool for feedback systems.

Discussion

Hanoch Ur (Polytechnic Institute of Brooklyn, Brooklyn, N. Y.): The problem of pole and zero sensitivity, as part of the problem of sensitivity in general, has been of interest in the network group at Polytechnic Institute of Brooklyn, and there has been work on it for some time. Some of the results shown by Mr. Huang have been given in papers by Papoulis,[1] Truxal and Horowitz,[2] and Angelo.[3] The definition of S_x^p was somewhat different in these other papers:

$$S_x^{p_i}=\frac{\partial p_i}{\partial \ln x}=\frac{\partial p_i}{\partial x}\frac{\partial p_i}{x} \tag{20}$$

No division by p_i was made. It leads to a dimensional $S_x^{p_i}$ which is apparently undesirable. This difficulty, however, turns out to be an advantage, due to the following theorem.

Using Mr. Huang's notation, consider the following:

$$H(s)=\frac{K\prod_{i=1}^{m}(s+z_i)}{\prod_{i=1}^{n}(s+p_i)} \tag{21}$$

and take the total logarithmic derivative, assuming that K and all z_i and p_i are functions of a single parameter x.

$$\frac{dH}{H}=dx\left(\frac{\frac{\partial K}{\partial x}}{K}+\sum_{i=1}^{m}\frac{\frac{\partial z_i}{\partial x}}{s+z_i}+\sum_{i=1}^{n}\frac{-\frac{\partial p_i}{\partial x}}{s+p_i}\right) \tag{22}$$

If the expression is now divided by $\frac{dx}{x}$, there is obtained

$$S_x^H=\frac{\frac{dH}{H}}{\frac{dx}{x}}=\frac{\frac{\partial K}{K}}{\frac{\partial x}{x}}+\sum_{i=1}^{m}\frac{\frac{\partial z_i}{\partial x}/x}{s+z_i}+\sum_{i=1}^{n}\frac{\frac{\partial(-p_i)}{\partial x}/x}{s+p_i} \tag{23}$$

The expression on the left is the well-known sensitivity function in the complex frequency s. It is shown, on the right hand side, as a partial function expansion, in which all critical frequencies become poles and the residues are the pole-zero sensitivities as defined here.

This theorem, which interrelates the apparently independent definitions of sensitivity, is interesting and is perhaps necessary to a proper perspective.

The proof of the main theorems (equations 8 through 10) can be made more elegant. An example is the evaluation of S_K^p.

$$G(s)=\frac{KP(s)}{Q(s)} \tag{24}$$

where P and Q are polynomials. The poles of $H(s)$ are the zeros of

$$F(s,k)\triangleq Q(s)+KP(s)=0 \tag{25}$$

This equation can be viewed[4] as an implicit relation between s and K. Therefore, if this expression is multiplied by

K, it is noted that the expression on the right is

$$\frac{\dfrac{\partial F}{\partial K}}{\dfrac{\partial F}{\partial s}} = -\frac{P(s)}{\dfrac{\partial}{\partial s}[P(s)+KQ(s)]} \quad (26)$$

which is the well-known form for the residue of H

$$\frac{\partial s}{\partial K/K} = -\frac{KP(s)}{\dfrac{\partial}{\partial s}[P(s)+KQ(s)]}$$

$$= -\text{res}\, H(s,K) = -k_p \quad (27)$$

The same methods also were used advantageously to obtain other sensitivity expressions. The interpretation of the sensitivity equations leads to new compensation techniques which are, however, outside the scope of the paper discussed.

It might be proper to mention again the importance of the sensitivity concept in feedback networks and active networks in general. In the opinion of this writer it is always important to go back to the original definition of the sensitivity function used, and to understand its meaning in a given case. To this end various sensitivity functions may be used. It is, however, of utmost importance to find the relations between the numerous definitions. These relationships also imply general limitations on compensation techniques.

References

1. Displacement of the Zeros of the Impedance $Z(p)$ Due to Incremental Variations in the Network Elements, A. Papoulis. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Jan. 1955, p. 79.

2. Sensitivity Considerations in Active Network Synthesis, J. G. Truxal, I. Horowitz. *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, vol. CT-3, no. 4, Dec. 1956.

3. Design of Feedback Systems, E. J. Angelo, Jr. *Research Report R-449-55, PIB-379*, Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1955.

4. Geometrical Properties of Root-Loci, H. Ur. *M. S. Thesis*, Purdue University, Lafayette, Ind., Aug. 1956.

R. Y. Huang: The author wishes to thank Mr. Ur for his helpful and interesting comments. In regard to the definition of $S_x^{p_i}$ as

$$S_x^{p_i} = \frac{\partial p_i}{\partial x}\bigg|_x$$

which Mr. Ur considers an advantage because of the form of S_x^H, it seems to the author that the advantage is a purely formal one. The undesirability of defining $S_x^{p_i}$ and S_x^H in different forms, i.e., S_x^H defined as

$$S_x^H = \frac{\dfrac{dH}{H}}{\dfrac{dx}{x}}$$

for the purpose of a purely formal simplicity, seems to offset the advantage of this simplicity. It would not seem that much simplicity is lost by defining $S_x^{p_i}$ in the usual manner, in which the residues at the critical frequencies become the pole-zero sensitivities, multiplied by the pole-zero positions.

The proofs of $S\sigma_i^p$ and $S\omega_i^p$, made in the manner proposed, would require a slight complication not mentioned by Mr. Ur. This complication arises from the fact that $S\sigma_i^p$ and $S\omega_i^p$ were found by maintaining the zero-frequency gain constant. These particular definitions also result in the difficulty that the general sensitivity function S_x^H as given by Mr. Ur could not be applied directly to the sensitivities $S\sigma_i^p$ and $S\omega_i^p$.

The general sensitivity function S_x^H is certainly of great interest. It enables one to describe the sensitivity of $H(s)$ to x at all complex frequencies s. Of course, in the case of systems where there are only two dominant time constants; or where the system possesses two distinct and well-separated modes of response, it would be sufficient, for all practical purposes, to find the sensitivities of these "control poles" to the given parameter.

In conclusion, the author is gratified to learn that work similar to his own is being done elsewhere; it seems that, on the complex frequency plane as in many other places, the application of the concept of linearizing in the small promises to yield some results.

Use of Phase Space in Transient-Stability Studies

S. T. BOW
ASSOCIATE MEMBER AIEE

J. E. VAN NESS
MEMBER AIEE

IN THE PAST FEW YEARS a great deal of interest has developed in the analysis of nonlinear systems. Much of this has been the result of the great strides taken in the field of control systems. However, nonlinearities have a way of appearing in many forms in many different types of systems. In fact, almost all systems when analyzed carefully have some degree of nonlinearity, and the linear representations are only approximations. In this paper some of the techniques that originated in control system work are extended to the analysis of a power system's stability, and in the process some new methods of analyzing nonlinear systems are presented.

The methods used here describe the response of the third-order differential equation of the system in terms of a 3-dimensional phase space. Two approaches are presented. The first gives an indication of the response of the system through the use of sketches of the cross sections of the trajectories in the phase space. The second makes use of surfaces in the phase space which form the stability boundaries of the system. While these methods are applied to a problem arising in the study of the stability of a power system, it is hoped that they will be helpful in other systems which give rise to third-order nonlinear equations.

The use of the phase plane for second-order systems has been presented many times.[1-2] The first method presented in this paper, the sketching of the intercepts of the trajectories, is analogous to the method of isoclines in the phase plane, while the stability surfaces bear a relationship to a separatrix in the phase plane. The use of a space with more than two dimensions for higher order systems has been suggested often in the literature on nonlinear systems. Two of the more modern applications of a 3-dimensional space using the same coordinates used in this paper are by Ku[4] and Hopkin.[5]

While two examples are given to illustrate the methods presented, the intention of this paper is to present methods, not to analyze any specific system. In fact, more work will have to be done before the true value of these methods can be evaluated, but they do show a great deal of promise as another tool in evaluating the stability of a nonlinear system.

Phase Space

In general, the condition of a system at any instant of time can be described by giving the values of certain variables of the system at that instant. The number

Paper 58-83, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 22, 1956; made available for printing December 10, 1957.

S. T. Bow is with the Electrical Engineering Research Institute, Chinese Academy of Science, Nanlin, Chang Chun, China, and J. E. Van Ness is with Northwestern University, Evanston, Ill.

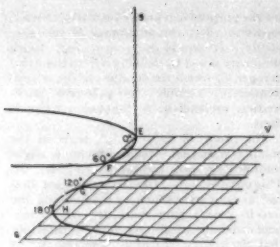


Fig. 1. Sketch of the singular points in v-δ plane

of variables required depends on the order of the equations used to describe the system. Thus, a simple pendulum, which can be represented by a second-order differential equation, is described at any instant by giving its position and velocity at that instant. In this paper, the authors have used as an example a third order nonlinear system which can be described in terms of several different variables, the power angle, field current, terminal voltage, etc. For convenience in analyzing the problem and in interpreting the results, this system is described in terms of the power angle δ, and its first and second derivatives with respect to time, v and g.

These three variables can be used to describe a space, called the phase space, as shown in Fig. 1. The plane in which the g and v axes lie can be thought of as being in the plane of the page, while the δ axis is perpendicular to the page. The motion of the operating point of the system in this space is in part dictated by the definitions of g and v. That is, whenever the operating point is in the right-hand half space where v is positive, δ must be increasing in value. In the left-hand half space, δ must be decreasing in value. Likewise, in the upper half space where g is positive, the operating point must be moving toward the right to a higher value of v, and in the lower half space it must be moving to the left. Thus, an operating point in the upper right-hand front octant of the space must be moving to the right and out of the page.

The motion of the system under any operating condition can be described by a curve in this space; or the general operating characteristics of the system can be studied if sketches of trajectories or flow lines for all of the space of interest can be made. Stationary points representing stable operating points can only exist on the δ axis where g and v are zero. If the system is to be stable, the trajectories must tend to spiral in and converge on these points. Because of the difficulty in sketching a family of 3-dimensional curves, a point that appears to be a stable operating point must be tested by other means to be sure that it is truly stable. However, the sketches will definitely show many areas of the δ axis that cannot be stable.

In linear systems the stability of the system is independent of the boundary values or initial conditions. Therefore, with a linear system the whole space represents either a stable or unstable condition. However, in nonlinear systems the stability of the system is often dependent on the initial conditions. When this is true, certain regions of the phase space will lead to stable operation, either a stable operating point on the δ axis, or a stable limit cycle representing a sustained oscillation, while other regions will lead to instability. The surface dividing these two regions is referred to as the stability boundary surface. Two similar systems can be compared by comparing the shape of their stability boundary surfaces. Thus, a system which had a larger volume enclosed by this surface might be considered more stable than one whose surface enclosed a smaller volume. Of course, the shape of the surface and the part of the space enclosed would also be important.

Method of Approach

Sketching the Trajectories

The purpose in sketching the trajectories of the system is to make use of topological methods which will give a qualitative picture of the behavior of the system. However, it is very difficult to sketch 3-dimensional curves. Therefore, cross sections are taken through the space, and the direction and curvature of the trajectories where they pass through these cross sections are indicated on the resulting plane. A group of these cross sections formed at regular intervals through the space being studied will give a picture of the motion of the system through that space. It should be emphasized that these sketches show the intersections of the trajectories with the plane chosen, not the projections of the trajectories on the plane. Many choices could be made as to the planes to use for the cross sections, but the planes formed by setting δ equal to a set of constants seem the most satisfactory.

The first step in sketching the trajectories is to transform the equation of the system into the form

$$\frac{dg}{dv} = f(g, v, \delta) \tag{1}$$

where

$$g = \frac{dv}{dt} \tag{2}$$

This usually can be done without too much difficulty as long as the original equation does not contain t explicitly, i.e., the differential equation describing the system should not involve time-varying parameters. If it does, phase-space methods cannot be used for the same reasons that the phase plane cannot be used in a second-order system containing t explicitly. The differential dt can be eliminated by use of equation 2. The right-hand side of equation 1 will usually be in the form of a fraction with g in the denominator.

In a manner similar to that used with 2-dimensional phase planes, the points in the phase space can be classified as singular points and ordinary points. The singular points are of interest because they may represent stable operating points. They can be found by solving equation 1 for those values of g, v, and δ which make dg/dv indeterminate. Any of the usual methods of determining stability in the neighborhood of a point can be used to test the singular points to find whether they are stable or unstable.

By substituting fixed values for δ into equation 1, it is reduced to an equation with only two variables, g and v. The equation can then be used to sketch the behavior of the trajectories as they pass through the g-v plane having the chosen value of δ. This is illustrated in the example given later.

Stability Boundary Surfaces

Since at any instant the condition of the system can be described completely by giving the values of g, v, and δ, any operating condition can be represented by a point in the phase space. The motion of the system away from this point is described by the trajectories which are discussed in the previous section. Except at the singular points, these trajectories are unique so that from any given operating point the future motion of the system can be determined. If the question of interest is whether the system is stable or not with certain initial values, the answer can be given by describing those regions of space from which trajectories lead to a stable operating point and those from which they do not. The dividing surface between these regions will be referred to as the stability boundary surface.

So far, no method has been discovered

which will analytically give this surface. However, the surface can be found rather easily using an analog computer. The system is set up on the computer and then different initial conditions tested to see whether they lead to stable or unstable operation. By systematically choosing the conditions tried, the surface can be found without too much difficulty.

Example Problem

The method of applying the phase-space concept to a nonlinear system can best be explained with an example. For this paper the example chosen was a cylindrical rotor synchronous machine connected through a reactance to an infinite bus. In representing the machine, the effect of changing flux linkages of the field has been included, but the machine has been assumed to have a cylindrical rotor, and the effects of the damper bars have been neglected. Thus the representation goes one step further than the conventional representation used in transient-stability studies in that the effect of changing flux linkage has been included. The system is described by equation 3. In this equation, δ is the angle of the rotor of the machine with respect to the voltage of the infinite bus.

$$A\frac{d^2\delta}{dt^2}+\frac{d^2\delta}{dt^2}-A\cot\delta\frac{d^2\delta}{dt^2}\frac{d\delta}{dt}+\left(\frac{P_t}{M}A\cot\delta+C\sin^2\delta\right)\frac{d\delta}{dt}+B\sin\delta-\frac{P_t}{M}=0 \quad (3)$$

A, B, C, P_t, and M are constants of the system. Since this paper is more interested in the method of analyzing equation 3 than in the particular system which it represents, the definitions and values used for these constants are omitted. These values and the derivation of equation 3 are given in the dissertation[4] on which this paper is based.

No direct method of obtaining an analytical solution of this equation seems to be available. In order to get equation 3 into a more convenient form for phase-space analysis, g and v are substituted to give equation 4. Time no longer appears explicitly in this equation, so that it can be used to find the trajectories of the system in phase space.

$$\frac{dg}{dv}=\frac{\left(\frac{P_t}{AM}-\frac{B}{A}\sin\delta\right)-\left(\frac{P_t}{M}\cot\delta+\frac{C}{A}\sin^2\delta\right)\frac{g}{A}+gv\cot\delta}{g} \quad (4)$$

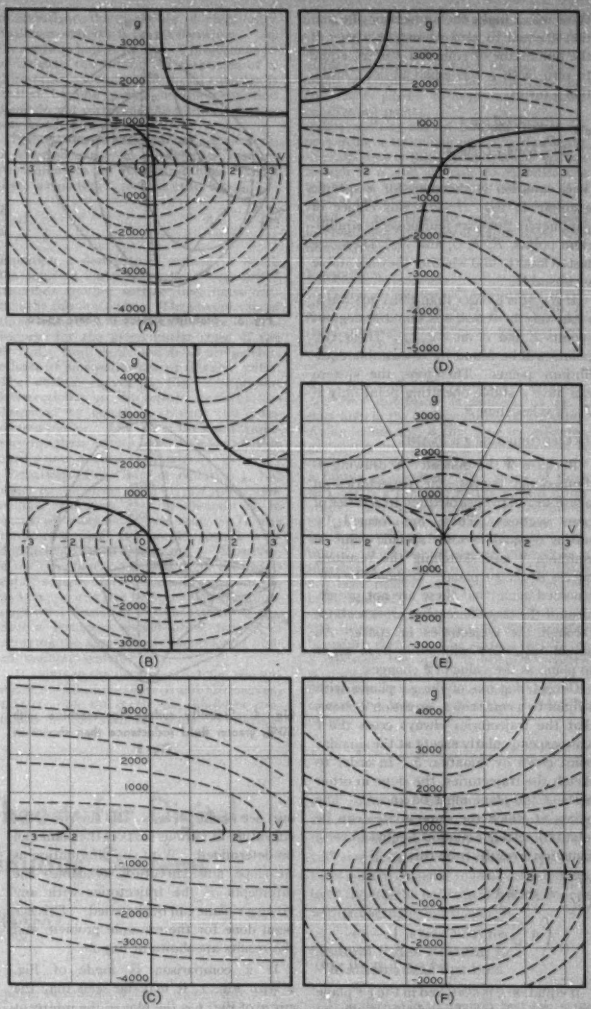


Fig. 2. g-v intercepts of trajectories in phase space

A—δ = 30 degrees
B—δ = 60 degrees
C—δ = 90 degrees
D—δ = 120 degrees
E—δ is slightly less than 0 or 180 degrees
F—δ is slightly greater than 0 or 180 degrees

The Singular Points

The first question in examining a given system is whether there can be any stable operating points. As was mentioned before, all stable operating points must lie on the δ axis in phase space where g and v are zero. When g is set equal to zero in equation 4 it is seen that dg/dv will be infinite except when the numerator is also equal to zero. The points of interest are

those which make the numerator of equation 4 equal to zero when g is zero. If the numerator of equation 4 is set equal to zero, with g also equal to zero, equation 5 result.

$$v = \frac{P_t/M - B \sin \delta}{A \frac{P_t}{M} \cot \delta + C \sin^2 \delta} \qquad (5)$$

This equation is sketched for a specific example in Fig. 1. For this choice of parameters there are four possible equilibrium points, E, F, G, and H. These are tested later to find whether they are stable or unstable.

If P_t/M is greater than B in equation 5, v can only be zero if δ is 0 or 180 degrees (points E and H on Fig. 1). These two points will be found to be unstable equilibrium points. Therefore, the system can have a stable operating point only if P_t/M is less than B.

Trajectories in g-v Plane

Because of the difficulty in drawing 3-dimensional curves, the trajectories in phase space are best pictured by a series of cross sections. These are formed by setting δ equal to some specific value in equation 4 and sketching the resulting curves in the g-v plane. It should be emphasized again that these are not projections of the trajectories, but cross sections through the trajectories in space. An actual trajectory will move from plane to plane as the value of δ changes.

Considering one of the g-v planes with δ equal to a constant δ_1, equation 4 shows that the trajectories always cross the v axis perpendicularly except at the singular point given by equation 5. In order to sketch the trajectories, the slope in other parts of the plane must be known. The points at which the slope is zero can be found by setting equation 4 equal to zero. Equation 6 results from this.

$$\left(\frac{P_t}{M} - B \sin \delta_1\right) - \left(A \frac{P_t}{M} \cot \delta_1 + C \sin^2 \delta_1\right) v - g + (A \cot \delta_1) gv = 0 \qquad (6)$$

If equation 6 is sketched in the g-v plane for a given δ, two hyperbolas result, as shown in Fig. 2(A). Since these curves represent the place where dg/dv is zero, the trajectories will cross these curves horizontally. In addition, on one side of the curves the slopes will all be positive, and on the other side they will all be negative. Whether the slope is positive or negative in any given region can be quickly evaluated from equation 4.

In a similar manner the second derivative of g with respect to v can be formed and set equal to zero. In this way the curvature in various parts of the plane can be determined. With the information as to slope and curvature available, the intercepts of the trajectories with any given g-v plane can be sketched. This has been done for the example problem and the results are shown in Fig. 2.

If a comparison is made of Fig. 2 with Fig. 1, it may be seen that the curve of Fig. 1 is the plot of the intercept of the zero-slope line with the v-δ plane. If Figs. 2(A) through (G) are pictured as being lined up in front of each other along the δ axis, some picture of the trajectories in phase space can be visualized.

Note that two sketches are given for 0- and 180-degree points. While it may seem that there is a discontinuity at these points, there actually is not. The location of the line of zero slope is the same in both figures. The difference comes from the changing of the location of the line for zero curvature. In Fig. 2(E) the lines of zero curvature form an X through the origin. As δ approaches 0 or 180 degrees, these lines move toward the g axis. At the transition point between Figs. 2(E) and (F) where δ is exactly equal to 0 or 180 degrees, the slope of the trajectories in all parts of the plane changes sign due to the change in sign of the cot δ term of equation 4. When these changes are considered, it is seen that there is a logical transition from Fig. 2(E) to Fig. 2(F).

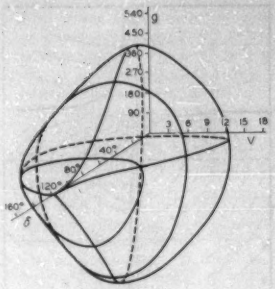


Fig. 3. Stability surface in phase space

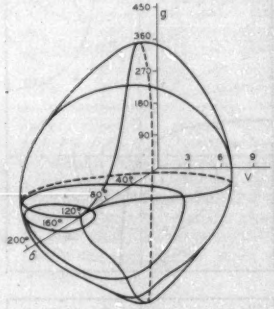


Fig. 4. Stability surface for machine with 50% greater field inductance than shown in Fig. 3

Stability of the Singular Points

For the values of the parameters chosen in the foregoing example, the singular points shown in Fig. 1 fall at 0, 60, 120, and 180 degrees. Comparing these to the corresponding trajectory sketches in Fig. 2, it appears that only the 60-degree point would be stable. To check this equation 1 can be linearized in the region of a singular point and the stability checked by the Hurwitz criterion. When this is done,[6] it is found that stable operating points can exist only for steady-state values of δ between $+90$ degrees and -90 degrees, with the specific exception of the origin, δ equal to zero. Therefore, for the example being considered here, only the 60-degree singular point will be stable. It should be recognized that equation 3 is periodic in δ, so that other stable points will occur at 420 degrees, 780 degrees, etc., but only the principal values of δ are being considered in this study.

Stability Boundary Surfaces

The surface for this system was found on the analog computer and is shown in Fig. 3. Any points which lie within the surface will lead to a stable operating point, while those that are outside will be unstable. While only one surface is shown, it should be remembered that the equation for this system is periodic in δ, so that there will be a series of these stable region along the δ axis.

Another stability surface is shown for comparison in Fig. 4. For this case the inductance of the field of the machine was increased by 50%. By comparing Fig. 4 with Fig. 3 it is seen that the area enclosed has been reduced. Thus, as a general statement, it could be said that the stability of the system has been reduced. However, for a more exact evaluation of the stability, the sections of each surface should be drawn and compared. A total increase in volume may not be as important as the regions that are increased or decreased.

Conclusions

The methods presented in this paper should be valuable in the analysis of third-order nonlinear systems. Their value lies chiefly in helping to form a pictorial representation of the system's response. In this sense their role is similar to that of equal-area criteria used in power-system stability problems. By their use the engineer may be able to see more easily the nature of the system with which he is working.

References

1. NON-LINEAR VIBRATIONS (book), J. J. Stoker. Interscience Publishers, Inc., New York, N. Y., 1950.

2. NON-LINEAR MECHANICS (book), N. Minorsky. J. W. Edwards Brothers, Ann Arbor, Mich., 1947.

3. A PHASE-PLANE APPROACH TO THE COMPENSATION OF SATURATING SERVOMECHANISMS, Arthur M. Hopkin. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 631–39.

4. A METHOD FOR SOLVING THIRD ORDER AND HIGHER ORDER NON-LINEAR DIFFERENTIAL EQUATIONS, T. H. Ku. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 256, no. 3, Sept. 1953, pp. 229–44.

5. A STUDY OF A PREDICTOR-TYPE AIR-FRAME CONTROLLER DESIGNED BY PHASE-SPACE ANALYSIS, A. M. Hopkin, Morimi Iwama. *AIEE Transactions*, vol. 75, pt. II, Mar. 1956, pp. 1–9.

6. USE OF PHASE SPACE IN TRANSIENT STABILITY STUDIES, S. T. Bow. Dissertation submitted to Northwestern University, 1956, available from University Microfilms, Ann Arbor, Mich.

Discussion

Frederick W. Nesline, Jr. (Massachusetts Institute of Technology, Cambridge, Mass.): The idea of using phase-space cross sections to investigate the responses of third-order nonlinear systems is an excellent one. As the authors imply, the use of projections to investigate the totality of responses of a third-order nonlinear differential equation is limited because points in the projection plane do not correspond to a unique trajectory in the phase space.

It should be emphasized that a trajectory, or solution curve, in phase space requires two equations for its description. One of these is equation 1, and the other is

$$\frac{dv}{d\delta} = \frac{g}{v} \tag{7}$$

Equation 1 describes a surface in which the solution curve must lie. Equation 7 describes another surface that must contain the solution curve. The intersection of these surfaces is the solution curve, or trajectory, for the given initial state of the system. Both dg/dv and $dv/d\delta$ are indeterminate at an equilibrium, or singular, point. The dashed curves in Fig. 2 represent intersections of trajectory surfaces (as described by equation 1) with the cross-section planes. The use of the word trajectory for these curves as well as for solution curves may confuse the reader.

Equations of the form of equation 1, but which contain time explicitly, i.e., nonautonomous equations, may be thought of in terms of extended phase space.[1] In the third-order case, this would require a 4-dimensional space δ, v, g, and t. However, if the time function is quantized, the differential equation is sequentially autonomous, and trajectories can be developed on a time-sequential basis in 3-dimensional space. Such a technique should prove helpful in studying the approximate behavior of nonautonomous systems.

It seems to the discusser that the two unstable equilibrium points (0 and 120 degrees) that straddle the stable one should be contained in the stability boundary surface, but in Figs. 3 and 4 the 0-degree point does not appear to be included. If these points should be a part of the stability boundary surface, an analytic solution for this surface might be possible since this is the only surface that passes through both unstable equilibrium points.

The authors should be commended for presenting a technique useful in describing the responses of third-order systems. Even though it is still difficult to visualize the totality of trajectories, one wonders if a similar technique, applied algebraically instead of geometrically, might be helpful in visualizing responses of higher order nonlinear systems.

REFERENCE

1. OPTIMUM RESPONSE OF DISCONTINUOUS FEEDBACK CONTROL SYSTEMS, F. W. Nesline, Jr. "Proceedings of the Computers in Control Systems Conference," *AIEE Special Publication T101*, May 1958, pp. 169-76.

S. T. Bow and J. E. Van Ness: Professor Nesline's discussion is appreciated as it may help in the understanding of the paper. As he points out, if the three co-ordinates, g, v and δ, are considered as independent of each other, then equation 1 does describe a family of surfaces. The shape of these surfaces is a function of the system under study. The actual trajectories are formed by the intersection of these surfaces with another set of surfaces described by equation 7. This second set of surfaces is a function of the manner in which the space has been defined, and not of any system being studied.

Only the unstable equilibrium point at 120 degrees is included in the stability boundary surface. The surface crosses the δ axis near the origin at about 8 degrees. A similar situation sometimes occurs in a 2-dimensional phase plane when a seperatrix crosses the axis at other than a singular point.

Statistical Design Theory for Digital-Controlled Continuous Systems

S. S. L. CHANG
MEMBER AIEE

IN A PREVIOUS PAPER,[1] the basic theory and a procedure for optimum design of strictly digital systems were presented. In practice, however, strictly digital systems are few in number. Of far greater interest are digitally controlled continuous systems, in which a continuous plant is controlled by intermittently monitored and processed error signals. While only error signals at specific instants are used to control the system, the mean-square error averaged over all times is to be kept at a minimum. This paper develops the basic theory into a design procedure for doing this. The design procedure is no more complicated than that for a strictly digital system.

In the past few years, considerable attention had been given to the problem of system error between sampling periods in a sampled-data feedback control system. Among the most notable contributions along this line were Jury's modified z-transform method[2] and Lago's method of hypothetically multiplying the sampling rate.[3] Both methods require considerable numerical manipulations even for the simplest systems. By minimizing the mean- or integral-square error at all times, the new design method makes it possible for a system designer to have some control over the error in between sampling periods without actually calculating it.

General Rules Relating to Mathematical Symbols

It is desirable to explain first the symbols which are used in the paper. Conventional rules are followed: Lower-case letters represent time functions. Capitalized letters represent either Laplace or

Paper 58-90, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted July 10, 1957; made available for printing November 12, 1957.

S. S. L. CHANG is with New York University, New York, N. Y.

z-transforms depending on whether the ensuing transform variable is s or z. Starred capital letters represent sampled transforms: e.g., $E(s)$ is the Laplace transform of the system error $e(t)$; $E^*(s)$ is the Laplace transform of sampled $e(t)$; $E(z)$ is the z-transform of sampled $e(t)$, and is obtained by substituting z for e^{Ts} in $E^*(s)$.

Substitutions of transform variables are indicated by writing $-s$ instead of s, or z^{-1} instead of z as the independent variable. For instance, $G(-s)$ is obtained by substituting $-s$, for s in $G(s)$, and $G(z^{-1})$ is obtained by substituting z^{-1} for z in $G(z)$.

The z-transform for $A(s)\,B(s)$ is written as $AB(z)$. However, the author is not aware of any accepted way of writing the z-transform of $A(-s)\,B(s)$. It is denoted therefore as $\hat{A}B(z)$ with the symbol $\wedge$ indicating that s is replaced by $-s$ in the capped function before the conversion to z-transform is made.

Nomenclature

$c(t)$ = controlled variable, or actual system output
$c_1(t)$ = desired output signal
$e(t)$ = system error, $c_1(t)-c(t)$
$e_2(t)$ = forcing function, applied directly to plant
$n(t)$ = noise signal, mixed up with reference input
$r(t)$ = reference input
$r_1(t)$ = total input, $r_1(t)=r(t)+n(t)$
$C(s)$, $C_1(s)$, $E(s)$, $E_2(s)$, $N(s)$, $R(s)$, $R_1(s)$ = Laplace transforms of respective signals
$C(z)$, $C_1(z)$, $E(z)$, $E_2(z)$, $N(z)$, $R(z)$, $R_1(z)$ = z-transforms of respective signals
$D(z)$ = transfer function describing sampled-data processing unit
$F(z)$ = undetermined factor of system function
$G(s)$, $G(z)$ = transfer function describing the plant and the hold circuit
$K(z)$ = system function $C(z)/R_1(z)$
$L(s)$, $L(z)$ = linear operator relating $c_1(t)$ to $r(t)$
$P(z)$ = product of factors in numerator of $G(z)$ with zeroes outside the unit circle (equation 51)
$Q(z)$ = meromorphic function of z, to be separated into two terms, $Q_1(z)$ and $Q_0(z)$
$Q_1(z)$, $Q_0(z)$ = terms in partial fraction expansion of $Q(z)$ with their poles inside and outside the unit circle respectively
$Y(z)$ = product of factors of $\hat{G}G(z)/G(z)G(z^{-1})$ with poles and zeroes inside the unit circle (equation 5)
$\chi(z)$ = arbitrary function of z with all its poles outside the unit circle
C, C_1, D, etc. = short notations for $C(z)$, $C_1(z)$, $D(z)$, etc., respectively
C^+, C_1^+, D^+, etc. = short notations for $C(z^{-1})$, $C_1(z^{-1})$, $D(z^{-1})$, etc., respectively
$\hat{C}$, $\hat{C}_1$, $\hat{G}$, etc. = short notations for $C(-s)$, $C_1(-s)$, $G(-s)$, etc., respectively
m'' = number of sampling periods which elapses before the first response of the plant to a step input is noted at a sampling instant
m' = constant, $m''+l$, step 2
T = sampling period
$\varphi_{11}(n,T)$ = autocorrelation sequence of signal 1
$\varphi_{12}(n,T)$ = crosscorrelation sequence of signals 1 and 2
$\Phi_{11}(z)$ = autocorrelation series of signal 1
$\Phi_{12}(z)$ = crosscorrelation series of signals 1 and 2
$cl\{Q(z)\}$ = constant term in double-ended power-series expansion of $Q(z)$
$\varphi_{11}(\tau)$ = autocorrelation function of signal 1, also $\varphi_{11}(t)$
$\varphi_{12}(\tau)$ = crosscorrelation function of signals 1 and 2, also $\varphi_{12}(t)$
$\Phi_{11}(j\omega)$ = spectral density of signal 1, also $\Phi_{11}(s)$
$\Phi_{12}(j\omega)$ = cross-spectral density of signals 1 and 2, also $\Phi_{12}(s)$

Optimum Design Procedure

Two types of optimum design problems are treated in this paper, namely:

1. Systems with least-integral-square error following an input of known shape.
2. Systems with least-mean-square error for stationary random inputs.

The first case is referred to as design for optimum transient response and the second case as design for optimum response to random inputs. In the second case, noise and prediction are among the factors considered. Sometimes it is desirable to restrict the mean-square amplitude of a certain system variable. A slightly modified procedure for accomplishing this additional requirement is also given.

Case 1. Design for Optimum Transient Response

Refer to Fig. 1 in which a continuous plant with known transfer function $G(s)$ is actuated by the sampled output $E_2^*(s)$ from the processing unit. Let $r(t)$ and $c(t)$ represent the reference input to the system and the output variable from the plant respectively, and $e(t)$ represent the system error $r(t)-c(t)$. The integral-square error following an input $r(t)$ beginning at $t=0$ is defined as

$$J=\int_0^\infty [e(t)]^2 dt \tag{1}$$

The problem is to determine the impulse transfer function $D(z)$ of the processing unit such that J is a minimum. The plant $G(s)$ may have transportation lag, nonminimum phase characteristics, or may be unstable by itself.

The optimum design procedure is as follows:

1. Calculate the z-transform functions $G(z)$, $R(z)$, $\hat{G}G(z)$, and $\hat{G}R(z)$.

2. The zeroes, λ_i, of $G(z)$ outside the unit circle are grouped into a factor $P(z)$.

$$P(z)=\prod_{i=1}^{l}(1-z/\lambda_i) \tag{2}$$

Calculate $m'=m''+l$, where m'' is an integer representing the transportation lag of $G(z)$ in terms of sampling periods. Write $G(z)$ as

$$G(z)=z^{-m'}P(z)G_1(z) \tag{3}$$

where $G_1(z)$ represents the balance of the transfer function. All the zeros of $G_1(z)$ are inside the unit circle. Furthermore $G_1(z)$ is a constant as z approaches infinity.

3. Factorize $\hat{G}G(z)/G_1(z)G_1(z^{-1})$ into $Y(z)$ and $Y(z^{-1})$. $Y(z)$ contains all the critical frequencies inside the unit circle, while $Y(z^{-1})$ contains all the critical frequencies outside the unit circle. The poles and zeros of $\hat{G}G(z)/G_1(z)G_1(z^{-1})$ are in reciprocal pairs. Consequently, it can be written as

$$\frac{\hat{G}G(z)}{G_1(z)G_1(z^{-1})}=\frac{A^2\prod_i(1-\alpha_i z^{-1})(1-\alpha_i z)}{\prod_j(1-\beta_j z^{-1})(1-\beta_j z)} \tag{4}$$

where $|\alpha_i|\leqslant 1$ and $\beta_j\leqslant 1$. Write

$$Y(z)=\frac{A\prod_i(1-\alpha_i z^{-1})}{\prod_j(1-\beta_j z^{-1})} \tag{5}$$

By splitting the factors as prescribed in the foregoing, one makes certain that $Y(z)$ does not possess a pole or a zero at ∞. Similarly $Y(z^{-1})$ does not possess a pole or a zero at the origin.

4. Expand the function $Q(z)$ into partial fraction expansion

$$Q(z)=\frac{z^{-1}\hat{G}R(z)}{G_1(z^{-1})Y(z^{-1})}=\sum_i\frac{C_i}{z-\gamma_i}+\pi(z) \tag{6}$$

where the polynomial $\pi(z)$ may be zero. Collect terms with γ_i in and on the unit circle.

$$Q_1(z)=\sum_{|\gamma_i|\leq 1}\frac{C_i}{z-\gamma_i} \tag{7}$$

5. The optimum closed-loop system function is given as

$$K_0(z)=\frac{z^{-m'+1}P(z)Q_1(z)}{R(z)Y(z)} \tag{8}$$

6. The transfer function of the processing unit is

$$D(z)=\frac{K_0(z)}{G(z)[1-K_0(z)]} \tag{9}$$

Case 2. Design for Optimum Response to Random Inputs

Refer to Fig. 2 in which the desired output $c_1(t)$ is given as a linear operation $L(s)$ on the reference input $r(t)$, while the actual output $c(t)$ from the plant is due to the total input $r_1(t)$, which is the sum of $r(t)$ and noise $n(t)$. The system error is defined as

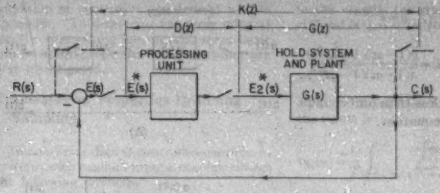


Fig. 1. Block diagram of a sampled-data system

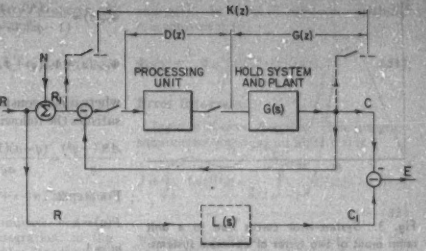


Fig. 2. Diagram defining system error E, which is to be minimized in the sense of least-mean-square amplitude

$e(t)=c_1(t)-c(t)$

The spectral densities $\Phi_{rir1}(j\omega)$ and $\Phi_{r1r}(j\omega)$ are known. The problem is to specify the impulse transfer function $D(z)$ to give least average value of $[e(t)]^2$.

The design procedure is as follows:

1. Replace $j\omega$ by s in the spectral density functions $\Phi_{rir1}(j\omega)$ and $\Phi_{r1r}(j\omega)$ to obtain $\Phi_{rir1}(s)$ and $\Phi_{r1r}(s)$ respectively. Calculate the z-transform functions $G(z)$, $\hat{G}G(z)$, $\Phi_{rir1}(z)$, and $\hat{G}L\Phi_{r1r}(z)$. If the spectral densities have been obtained by actual measurements, they are approximated as meromorphic functions of $j\omega$ before making the foregoing substitution.

2. Same as step 2 in case 1.

3. Split $\hat{G}G(z)/G_1(z)G_1(z^{-1})$ into $Y(z)$ and $Y(z^{-1})$ in the same way as step 3 in case 1. Similarly, split $\Phi_{rir1}(z)$ into $R_1(z)$ and $R_1(z^{-1})$. All the poles and zeros of $R_1(z)$ are inside or on the unit circle.

$$R_1(z)R_1(z^{-1})=\Phi_{rir1}(z) \tag{10}$$

4. Same as step 4 in case 1, except that $Q(z)$ is now defined as follows

$$Q(z)=\frac{z^{-1}\hat{G}L\Phi_{r1r}(z)}{G_1(z^{-1})Y(z^{-1})R_1(z^{-1})} \tag{11}$$

5. The optimum closed-loop system function is given as

$$K_0(z)=\frac{z^{-m'+1}P(z)Q_1(z)}{R_1(z)Y(z)} \tag{12}$$

6. Same as step 6 in case 1.

Systems with Constraint

In some systems, it is desirable to restrict the magnitude of a certain variable, say $v(t)$, in the plant. This can be done most conveniently by requiring the average value of $[v(t)]^2$ to be less than a certain limit M, or alternatively, the average value of $[v(t)]^2$ to be kept as low as possible while the mean-square error of the system is not to surpass a given limit. $G_c(z)$ is the transfer function $V(z)/E_0^*(z)$.

The foregoing condition of constraint can be incorporated into either design procedures as follows:

1. Instead of $Y(z)$ $Y(z^{-1})=\hat{G}G(z)/G_1(z)G_1(z^{-1})$, write

$$Y(z)Y(z^{-1})=\frac{\hat{G}G(z)+\eta\hat{G}_cG_c(z)}{G_1(z)G_1(z^{-1})} \tag{13}$$

where η is an undetermined positive constant. Increasing the value of η usually means decreasing the system gain.

2. Determine the optimum closed-loop system function $K_0(z)$ as before. The function $K_0(z)$ so determined has η as a parameter, and is denoted as $K_0(z)$.

3. Calculate the integral- or mean-square value of $v(t)$

$$\int_0^\infty [v(t)]^2dt=\frac{1}{2\pi j}\oint\frac{K_0(z)K_0(z^{-1})R(z)R(z^{-1})\hat{G}_cG_c(z)}{G(z)G(z^{-1})}\frac{dz}{z} \tag{14}$$

or

$$\overline{[v(t)]^2}=\frac{1}{2\pi j}\oint\frac{K_0(z)K_0(z^{-1})\Phi_{rir1}(z)\hat{G}_cG_c(z)}{TG(z)G(z^{-1})}\frac{dz}{z} \tag{15}$$

4. Determine the permissible range of η from the condition that the mean- or integral-square value of $v(t)$ is less or equal to M. The function $K(z)$ giving least-integral- or mean-square error with η in the permissible range is selected.

While steps 3 and 4 are usually tedious to perform, they may be skipped if η is to be determined by such considerations as gain or rise time rather than a rigid limit on the average value of $[v(t)]^2$.

Illustrative Examples

Example I. Illustrative Design for Least-Integral-Square Error with Known Input Function

With refe e ce to Fig. 1, let the transfer function of the t and the hold system be

$$G(s)=\frac{1-e^{-s}}{s}\times\frac{10}{s(1+s)^2}$$

The sampling time is assumed to be unity, $D(z)$ is to be designed to give least-integral-square error for ramp inputs, $R(s)=1/s^2$. Following the procedure outlined for case 1, the following results.

For step 1:

$$G(z)=\frac{1.02z^{-1}(1+2.34z^{-1})(1+0.16z^{-1})}{(1-z^{-1})(1-0.368z^{-1})^2}$$

$$R(z)=\frac{z^{-1}}{(1-z^{-1})^2}$$

$$\hat{G}G(z)=\frac{5.75(1+0.458z^{-1})(1+0.133z^{-1})}{(1+0.458z)(1+0.133z)}\cdot\frac{1}{(1-z^{-1})(1-z)(1-0.368z^{-1})^2(1-0.368z)^2}$$

For step 2:

$$G_1(z)=\frac{2.39(1+0.16z^{-1})}{(1-z^{-1})(1-0.368z^{-1})^2}$$

$$P(z)=1+0.427z$$

$$m'=2$$

For step 3:

$$Y(z)=\frac{(1+0.458z^{-1})(1+0.133z^{-1})}{(1+0.16z^{-1})}$$

For step 4, in determining $Q_1(z)$, only the relevant part of $\hat{G}R(z)$ needs to be calculated. There is a third-order root of $GR(z)$ at unity, while all the other roots of $Q(z)$ are outside the unit circle. Since the remaining factor has a single zero at unity, the relevant terms are the $1/(z-1)^3$ and $1/(z-1)^2$ terms. Therefore,

$$G(-s)R(z)=\frac{(1-e^s)10}{s^3(1-s)^2}\times\frac{1}{s^2}=10(1-e^s)\left[\frac{1}{s^4}+\frac{2}{s^3}\right]+\cdots$$

Correspondingly

$$\hat{G}R(z)=-10\left[\frac{1}{6}\frac{z+4z^2+z^3}{(z-1)^3}+\frac{z+z^2}{(z-1)^2}\right]+\cdots$$

$$Q_1(z)=\frac{1}{(z-1)^2}+\frac{2.403}{z-1}=\frac{2.403z^{-1}(1-0.584z^{-1})}{(1-z^{-1})^2}$$

For step 5:

$$K_0(z)=\frac{1.028z^{-1}(1+2.34z^{-1})(1-0.584z^{-1})(1+0.16z^{-1})}{(1+0.458z^{-1})(1+0.133z^{-1})}$$

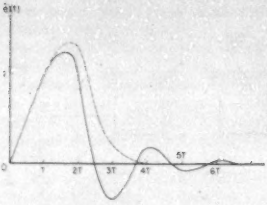


Fig. 3. System-error curves due to a unit ramp input of two types of optimum systems:

Solid line: system with least-integral-square error

Dashed line: deadbeat system

For step 6:

$$1-K_0(z)=\frac{(1-z^{-1})^2(1+1.404z^{-1})(1+0.159z^{-1})}{(1+0.458z^{-1})(1+0.133z^{-1})}$$

$$D(z)=\frac{K_0(z)}{G(z)[1-K_0(z)]}=\frac{1.028(1-0.584z^{-1})(1-0.368z^{-1})^2}{(1-z^{-1})(1+1.404z^{-1})(1+0.159z^{-1})}$$

The system error corresponding to a unit ramp input is plotted as the solid curve in Fig. 3. For comparison, the system error of an alternative system designed for absolute zero error after a few sampling periods is plotted as the dashed curve. (The method used for obtaining the alternative system is given in Appendix III.)

EXAMPLE II. ILLUSTRATIVE DESIGN FOR LEAST-MEAN-SQUARE ERROR WITH RANDOM INPUTS

With reference to Fig. 4, the known conditions are:

$$G(s)=\frac{1-e^{-sT}}{s}\times\frac{10}{1+2Ts}$$

$$L(s)=e^{nTs}$$

$$\Phi_{rr}(j\omega)=\frac{2\nu A^2}{\nu^2+\omega^2}$$

$$\Phi_{rn}(j\omega)=0$$

$$\phi_{nn}(\tau)=1,\ \tau=0$$

$$\phi_{nn}(\tau)=0,\ |\tau|\geq T$$

To design a data-processing unit which gives lowest mean-square error, the steps outlined for case 2 are followed.

For step 1:

$$G(z)=3.94z^{-1}/(1-0.606z^{-1})$$

$$\hat{G}G(z)=\frac{9.7T(1+0.264z)(1+0.264^{-1})}{(1-0.606z)(1-0.606z^{-1})}$$

$$\Phi_{nn}(z)=1$$

$$\Phi_{rr}(z)=\frac{A^2(1-q^2)}{(1-qz)(1-qz^{-1})},\quad q=e^{-\nu T}$$

$$\Phi_{r'r'}(z)=\Phi_{rr}(z)+\Phi_{nn}(z)=\frac{q(1-az)(1-az^{-1})}{a(1-qz)(1-qz^{-1})}$$

where a is a constant less than unity, and satisfies the following equation

$$\frac{A^2(1-q^2)}{q}=\frac{(q-a)(1-aq)}{aq}$$

For step 2:

$$P(z)=1$$

$$m'=1$$

For step 3:

$$Y(z)=2.5\sqrt{\frac{T_\varepsilon}{a}}(1+0.264z^{-1})(1-az^{-1})$$

For step 4:

$$G(-s)L(s)\Phi_{r'r'}(s)=\frac{20\nu A^2(1-e^{Ts})e^{nTs}}{(\nu^2-s^2)(1-2Ts)(-s)}=\frac{10A^2(1-q)q^n}{(\nu+s)(1+2T\nu)\nu}+\ldots$$

$$\hat{G}L\Phi_{r'r'}(z)=\frac{10A^2(1-q)q^{n+1}}{\nu(1+2T\nu)(z-q)}+\ldots$$

$$Q_t=\frac{10A^2(1-q)q^{n+1}}{\nu(1+2T\nu)[Y(z)G(z)]_{z^{-1}=q}}\times\frac{z^{-1}}{1-qz^{-1}}$$

For step 5:

$$K_0=\frac{Q_t}{Y}=\frac{1.264Cq^n(q-a)z^{-1}}{(1+0.264z^{-1})(1-az^{-1})}\qquad(16)$$

where

$$C=\frac{1-q}{T\nu q(1+2T\nu)}\times\frac{1.264}{(1+0.264q)}\times\frac{(1.606q)}{0.394}$$

The value of a in terms q and A^2 is given in Fig. 5 of reference 1. Compared with equation 61 of reference 1, equation 16 has an extra factor $(1+0.264z^{-1})$ in the denominator and a different proportionality constant. The factor $1.264/(1+0.264z^{-1})$ has the effect of speeding up the system response with a slightly increased response to noise.

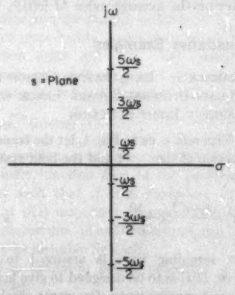


Fig. 4. Diagram showing division of jω axis into lengths of ω_s ($\omega_s=2\pi/T$)

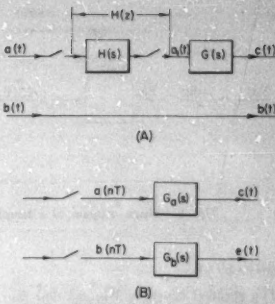


Fig. 5. Diagrams showing two typical cases of transmission through linear networks with sampling

Discussion

The foregoing optimum design procedures are based on the merits of the mean- (or integral) square-error criterion. For inputs of known time function, such as step or ramp, this criterion is at best a controversial one. It has a tendency of speeding up the system response at the cost of an oscillatory error tail. The design procedure is also somewhat more complicated than one which aims at a deadbeat system (zero error at all times after a few sampling periods).

However, for random inputs, absolute settling becomes less significant, since small inputs and disturbances are always present. Assuming that the successive inputs and disturbances are incoherent, the instantaneous system errors due to these inputs and disturbances add orthogonally to give the expected total instantaneous system error. Therefore, the mean-square error becomes the only logical criterion. Another viewpoint is that with ever-present disturbances, a faster response is more significant than absolute settling.

With multiple inputs, such as reference input, noise, and/or plant disturbance, the statistical correlations between various inputs must be taken into account. The present procedure represents the only way of doing it analytically.

Appendix I. Derivation of Optimization Procedures

The two basic procedures together with the modified procedure in the presence of a

condition of constraint are derived in this appendix. To determine the error integral for the case of random inputs, some basic transmission properties with sampling are first investigated.

Transmission Properties Involving Sampling

In analyzing digital-controlled continuous systems with random inputs, one encounters the following problems:

Problem 1. With reference to Fig. 5(A), if the correlation function $\varphi_{ba}(t)$ or the spectral density $\Phi_{ba}(s)$ is known, what is $\varphi_{bc}(t)$ or $\Phi_{bc}(s)$? Both source signals $a(t)$ and $b(t)$ are assumed to be ergodic. $H(z)$ and $G(s)$ may represent physical elements or hypothetical linear relations, such as prediction.

Let h_n and $g(t)$ represent the impulse responses of the respective elements. In terms of z-transform, it is

$$H(z)=\sum_{n=-\infty}^{\infty} h_n z^{-n} \tag{17}$$

It is understood that h_n vanishes for large negative values of n, and $g(t)$ vanishes for large negative values of t. The output variable $c(t)$ can be expressed as

$$c(t)=\sum_{n}\sum_{m} a(mT)h_{n-m}g(t-nT) \tag{18}$$

Equation 18 can be interpreted physically as follows: $\sum_m a(mT)h_{n-m}$ is the value of the impulse input to $G(s)$ at time nT. If $H(z)$ represents a physical element, it responds only to past and present inputs, then h_{n-m} vanishes for all values of $m>n$. If $H(z)$ represents prediction of k sampling intervals, h_{n-m} vanishes for all $m>n+k$. In either case, the summation can be taken for all values of m, from $-\infty$ to ∞. The output $c(t)$ is the sum total of the impulse response of $G(s)$ to all previous inputs. As before, the summation in n can be taken from $-\infty$ to ∞.

By definition of the correlation function

$$\varphi_{bc}(\tau)=\frac{1}{2T_1}\int_{-T_1}^{T_1} b(t)c(t+\tau)dt \tag{19}$$

where T_1 is a large interval approaching infinity. Let $b_1(t)$ represent the variable which is equal to $b(t)$ inside the interval $-T_1<t<T_1$ and zero elsewhere. The substitution of equation 18 into equation 19 gives

$$\varphi_{bc}(\tau)=\frac{1}{2T_1}\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\times$$
$$b_1(t)a(mT)h_n g(t+\tau-nT-mT)dt \tag{20}$$

Let $t'=t-mT$. Equation 20 can be written as

$$\varphi_{bc}(\tau)=\frac{1}{2T_1}\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\times$$
$$b_1(t'+mT)a(mT)h_n g(t'+\tau-nT)dt \tag{21}$$

In equation 21, the only factors depending on m are $b_1(t+mT)$ and $a(mT)$. Since b_1 is nonvanishing in an interval $2T_1$, there are, within a tolerance of ± 1, $2T_1/T$ nonvanishing terms in the summation over m. Due to the ergodic (or stationary) property of $a(t)$ and $b(t)$, one has for large T_1

$$\sum_{m=-\infty}^{\infty} b_1(t'+mT)a(mT)=\frac{2T_1}{T}\varphi_{ba}(-t') \tag{22}$$

By substituting equation 22 into equation 21 the result is

$$\varphi_{bc}(\tau)=\frac{1}{T}\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\times$$
$$\varphi_{ba}(-t')h_n g(t'+\tau-nT)dt' \tag{23}$$

Equation 23 represents composite convolution, summation, and integration, in the time domain. In terms of s-domain variables, it is

$$\Phi_{bc}(s)=\frac{1}{T}H^*(s)G(s)\Phi_{ba}(s) \tag{24}$$

Problem 2. With reference to Fig. 5(B), if $\Phi_{ab}(n,T)$ or $\Phi_{ab}(z)$ is known, what is $\Phi_{ca}(z)$ or $\Phi_{ca}(s)$? The same assumptions on $a(t)$, $b(t)$, $G_a(s)$, and $G_f(s)$ are made as in the previous case.

The output variable $c(t)$ can be considered as the sum total of impulse response of $G_a(s)$

$$c(t)=\sum_{n=-\infty}^{\infty} a(nT)g_a(t-nT) \tag{25}$$

A similar expression holds for $e(t)$. By definition

$$\varphi_{ce}(\tau)=\frac{1}{2T_1}\int_{-T_1}^{T_1}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\times$$
$$a(nT)b(mT)g_a(t-nT)g_b(t+\tau-mT)dt \tag{26}$$

As T_1 approaches infinity, the time interval in which $g_b(t)$ has appreciable value becomes negligible compared with T_1. Equation 26 can be written as

$$\varphi_{ce}(\tau)=\frac{1}{2T_1}\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\times$$
$$a(nT)b_1(mT)g_a(t-nT)g_b(t+\tau-mT)dt \tag{27}$$

where $b_1(nT)=b(mT)$ in the interval $-T_1<mT<T_1$ but zero elsewhere. Let $t'=t-nt$, and $m=n+m'$. Equation 27 becomes, by changing variables t and m to t' and m' respectively,

$$\varphi_{ce}(\tau)=\frac{1}{2T_1}\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\times$$
$$a(nT)b_1(nT+m'T)g_a(t')g_b(t'+\tau-m'T)dt' \tag{28}$$

Similar to equation 22, one derives

$$\sum_{n=-\infty}^{\infty} a(nT)b_1(nT+m'T)=\frac{2T_1}{T}\varphi_{ab}(m'T) \tag{29}$$

From equations 28 and 29

$$\varphi_{ce}(\tau)=\frac{1}{T}\int_{-\infty}^{\infty}\sum_{m'=-\infty}^{\infty}\times$$
$$\varphi_{ab}(m'T)g_a(t')g_b(t'+\tau-m'T)dt' \tag{30}$$

Equation 30 is the convolution sum and integral of three variables $\varphi_{ab}(m'T)$, $g_a(-t)$, and $g_b(t)$. In terms of s-domain variables, it is equivalent to

$$\Phi_{ce}(s)=\frac{1}{T}\Phi_{ab}{}^*(s)G_a(-s)G_b(s) \tag{31}$$

Error Integrals

Input of Known Function. The integral-square error following an input $r(t)$ is

$$J=\int_0^\infty [a(t)]^2dt=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} E(s)E(-s)ds \tag{32}$$

where $E(s)$ is the Laplace transform of the transient error $e(t)$. With reference to Fig. 1, $E(s)$ can be expressed as

$$E(s)=R(s)-C(s)$$
$$=R(s)-R^*(s)K^*(s)\frac{G(s)}{G^*(s)} \tag{33}$$

In equation 32, noise $N(s)$ is assumed to be zero.

To evaluate J, the path of integration is divided into segments of length ω_s as shown in Fig. 4. Let $\hat{E}(s)$ denote $E(-s)$. Equation 32 can be written as

$$J=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\hat{E}(s)E(s)ds=\frac{1}{2\pi j}\sum_{n=-\infty}^{\infty}\times$$
$$\int_{-\frac{j\omega_s}{2}}^{\frac{j\omega_s}{2}}\hat{E}(s+jn\omega_s)E(s+jn\omega_s)ds$$
$$=\frac{T}{2\pi j}\int_{-\frac{j\omega_s}{2}}^{\frac{j\omega_s}{2}}(\hat{E}E)^*(s)ds \tag{34}$$

The last equality sign is obtained by interchanging the order of summation and integration. It is justified on the physical basis that $|E(j\omega)|^2$ must converge rapidly to zero at least as fast as $1/\omega^2$ for large ω, and therefore the series $\sum_{n=-\infty}^{\infty}\hat{E}(s+jn\omega_s)\times E(s+jn\omega_s)$ converges uniformly. The same condition also ensures that the latter series is equal to $T(\hat{E}E)^*(s)$ exactly, without a constant corrective term.[4]

In terms of the variable z, equation 34 becomes

$$J=\frac{1}{2\pi j}\oint \hat{E}E(z)\frac{dz}{z} \tag{35}$$

The contour of integration is the unit circle. From equation 33, $\hat{E}E(z)$ can be evaluated as follows:

$$E(s)E(-s)=\left[R(s)-R^*(s)K^*(s)\frac{G(s)}{G^*(s)}\right]\times$$
$$\left[R(-s)-R^*(-s)K^*(-s)\frac{G(-s)}{G^*(-s)}\right]$$
$$=R(s)R(-s)+$$
$$\frac{R^*(s)R^*(-s)K^*(s)K^*(-s)}{G^*(s)G^*(-s)}\times$$
$$G(s)G(-s)-\frac{R^*(s)K^*(s)}{G^*(s)}G(s)R(-s)-$$
$$\frac{R^*(-s)K^*(-s)}{G^*(-s)}G(-s)R(s) \tag{36}$$

Apply z-transform to the foregoing equation, and note that $z^{-1}=e^{-sT}$. The result is

$$\hat{E}\hat{E}(z)=\hat{R}R(z)+\frac{R(z)R(z^{-1})K(z)K(z^{-1})}{G(z)G(z^{-1})}\times$$
$$\hat{G}G(z)-\frac{R(z)K(z)}{G(z)}\hat{G}R(z^{-1})-$$
$$\frac{R(z^{-1})K(z^{-1})}{G(z^{-1})}\hat{G}R(z) \quad (37)$$

Obviously, shorter notations for the z-transform functions are desirable. Let F denote $F(z)$, F^+ denote $F(z^{-1})$, and $<F_1F_2>$ denote $F_1F_2(z)$, etc. Equation 37 can be written then as

$$<\hat{E}\hat{E}>=<\hat{R}R>+\frac{RR^+KK^+}{GG^+}<\hat{G}G>-$$
$$\frac{RK}{G}<\hat{G}R>^+-\frac{R^+K^+}{G^+}<\hat{G}R> \quad (38)$$

Random Inputs. For random inputs, the mean-square error is

$$\bar{\epsilon}^2=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\Phi_{ee}(s)ds \quad (39)$$

Following the same steps which lead to equation 35, equation 39 becomes

$$\bar{\epsilon}^2=\frac{1}{2\pi j}\oint\Phi_{ee}(z)\frac{dz}{z} \quad (40)$$

Referring to Fig. 2, $e(t)$ is defined in general as the difference between the desired output $c_1(t)$ and the system output $c(t)$. The output functions can be expressed in terms of the input functions by means of Laplace transform

$$C(s)=\frac{R_i^*(s)K^*(s)G(s)}{G^*(s)} \quad (41)$$

$$C_1(s)=L(s)R(s) \quad (42)$$

$$E(s)=C_1(s)-C(s) \quad (43)$$

From equation 43, one obtains

$$\Phi_{ee}(s)=\Phi_{c_1c_1}(s)+\Phi_{cc}(s)-\Phi_{cc_1}(s)-\Phi_{cc_1}(s) \quad (44)$$

Equation 42 gives

$$\Phi_{c_1c_1}(s)=L(s)L(-s)\Phi_{rr}(s) \quad (45)$$

Equations 24, 41, and 42 give

$$\Phi_{cc}(s)=\frac{1}{T}\frac{K^*(s)G(s)}{G^*(s)}\Phi_{rr_1}(s)$$
$$=\frac{1}{T}\frac{K^*(s)G(s)L(-s)}{G^*(s)}\Phi_{rr_1}(s) \quad (46)$$

Applying z-transform to the foregoing equation, one obtains

$$\Phi_{cc}(z)=\frac{1}{T}\frac{K(z)}{G(z)}\hat{G}L\Phi_{rr_1}(z)$$
$$=\frac{1}{T}\frac{K(z)}{G(z)}\hat{G}L\Phi_{rr}(z^{-1}) \quad (47)$$

The second equality sign holds because $\Phi_{cc_1}(s)=\Phi_{cc}(-s)$, and $\epsilon^{-sT}=z^{-1}$. Similarly,

$$\Phi_{cc_1}(z)=\frac{1}{T}\frac{K(z^{-1})}{G(z^{-1})}\hat{G}L\Phi_{rr}(z) \quad (48)$$

From equations 31 and 41

$$\Phi_{cc}(z)=\frac{1}{T}\frac{K(z)K(z^{-1})G\hat{G}(z)\Phi_{rr_1}(z)}{G(z)G(z^{-1})} \quad (49)$$

Finally, from equations 44, 45, 47, 48, and 49, $\Phi_{ee}(z)$ is obtained and expressed in short notation:

$$\Phi_{ee}(z)=<\hat{L}L\Phi_{rr}>+\frac{1}{T}\frac{KK^+}{GG^+}<\hat{G}G>\Phi_{rr_1}-$$
$$\frac{1}{T}\frac{K}{G}<\hat{G}L\Phi_{rr}>^+-\frac{1}{T}\frac{K^+}{G^+}<\hat{G}L\Phi_{rr}> \quad (50)$$

Once $\hat{E}\hat{E}(z)$ and $\Phi_{ee}(z)$ are known, the system functions $K(z)$, which give least values of J and $\bar{\epsilon}^2$, can be determined.

Determination of Optimum System Function

The system function $K(z)$ must meet the following conditions of physical realizability:[1,5]

1. All the poles $K(z)$ are within the unit circle.
2. $K(z)$ must have at least as much transportation lag as the plant.
3. $K(z)$ must contain all the zeroes of $G(z)$ which are outside the unit circle.

The way of determining the optimum system function $K(z)$ is exactly the same as that for a strictly digital system as given in reference 1. The only difference is that equations 38 and 50 are now used instead of equation 30 of reference 1. Let

$$K(z)=z^{-m'}P(z)F(z) \quad (51)$$

where all the poles of the undetermined function $F(z)$ are inside the unit circle. Following the same procedure, one arrives at the conditions for the optimum system functions as listed in the following. For known input

$$\frac{z^{-1}RR^+PP^+<G\hat{G}^+>F_0}{GG^+}-\frac{z^{m'-1}R^+P^+}{G^+}$$
$$<\hat{G}R>=X(z) \quad (52)$$

For random inputs

$$\frac{z^{-1}PP^+<\hat{G}G>\Phi_{rr_1}F_0}{GG^+}-\frac{z^{m'-1}P^+}{G^+}$$
$$<\hat{G}L\Phi_{rr}>=X(z) \quad (53)$$

where $X(z)$ is an arbitrary function of z with all its poles outside the unit circle.

The solutions of equations 52 and 53 are given as follows: From equations 3 through 5, equation 52 can be written as

$$z^{-1}RR^+YY^+F_0-z^{-1}R^+<\hat{G}R>/G_1^+=$$
$$\chi(z) \quad (54)$$

Dividing equation 54 by R^+Y^+ yields

$$z^{-1}RYF_0-Q(z)=\chi(z)/R^+Y^+ \quad (55)$$

As all the poles of $z^{-1}\,RY\,F_0$ are inside the unit circle, and all the poles of $X(z)/R^+Y^+$ are outside the unit circle, equation 55 gives

$$z^{-1}RYF_0=Q_1(z) \quad (56)$$

The function Q_1 is defined in equation 7.

The optimum system function $K_0(z)$ is obtained from equations 51 and 56, and is expressed in equation 8. $D(z)$ is calculated from $K_0(z)$ as shown in equation 9.

For the case of random inputs, the derivation of the optimum design procedure is similar to the foregoing one.

Systems with Constraints

Similar to equation 34, the integral-square value of $v(t)$, following a transient input, can be expressed as

$$J_v=\frac{1}{2\pi j}\oint\hat{V}V(z)\frac{dz}{z} \quad (57)$$

For random inputs, the mean-square value of $v(t)$ is

$$\bar{v}^2=\frac{1}{2\pi j}\oint\Phi_{vv}(z)\frac{dz}{z} \quad (58)$$

By definition of $G_c(s)$, $V(s)$ can be expressed in terms of the input function as

$$V(s)=G_c(s)E_i^*(s)=\frac{G_c(s)K^*(s)R^*(s)}{G^*(s)} \quad (59)$$

From equation 59, one derives

$$\hat{V}V(z)=\frac{K(z)K(z^{-1})R(z)R(z^{-1})\hat{G}_cG_c(z)}{G(z)G(z^{-1})} \quad (60)$$

For the case with random inputs, the following is derived from equations 59 and 31

$$\Phi_{vv}(z)=\frac{1}{T}\frac{K(z)K(z^{-1})\Phi_{rr_1}(z)\hat{G}_cG_c(z)}{G(z)G(z^{-1})} \quad (61)$$

Equations 57, 58, 60, and 61 lead to equations 14 and 15.

There are two alternative ways in setting up the optimization criteria: 1. minimizing J or $\bar{\epsilon}^2$ while keeping J_v or $\bar{v}^2$ below a given limit, and 2. minimizing J_v or $\bar{v}^2$ while keeping J or $\bar{\epsilon}^2$ below a given limit. The first case represents designing for best performance while using a given fixed component. The second case represents designing for least cost to meet a given performance criterion.

In either case, Lagrange's method of undetermined multipliers is applied. The function to be minimized is $J+\eta J_v$ or $\bar{\epsilon}^2+\eta\bar{v}^2$, with η as the undetermined constant. This leads to equation 13, and the subsequent steps.

Appendix II. Proof of Equivalence of Alternative

Definitions of z-Transforms

Let $\Phi(s)$ be a real, rational function of s, with roots in both right-half plane and left-half plane, which converges at least as fast as s^{-2} as $s\to\infty$. In previous sections, $\Phi(z)$ is defined in three ways. The three differently defined $\Phi(z)$'s are denoted as $\Phi_a(z)$, $\Phi_b(z)$, and $\Phi_c(z)$ respectively. It is now shown that they are identical with each other.

The conventional definition, $\Phi_a(z)$, is

$$\Phi_a(z)=\sum_{n=0}^{\infty}\varphi_a(nT)z^{-n} \quad (62)$$

where

$$\varphi_0(t)=\frac{1}{2\pi j}\int_{\Gamma_1}\Phi(s)e^{st}ds \qquad (63)$$

and Γ_1 is the countour shown in Fig. 5. It consists of a vertical line aa' to the right of all the poles of $\Phi(s)$, and a large semicircle at infinity enclosing all the region to the left of aa'. Correspondingly, the series of equation 62 converges for values of z outside a circle on the z-plane with radius larger than unity. However, once expressed in closed form, $\Phi_0(s)$ also exists for values of z inside the circle. The tables for z-transform functions are usually calculated on this basis.

The correlation series is

$$\Phi_0(z)=\sum_{n=-\infty}^{\infty}\varphi_0(nT)z^{-n} \qquad (64)$$

where

$$\varphi_0(t)=\frac{1}{2\pi j}\int_{\Gamma_1}\Phi(s)e^{st}ds \quad \text{for } t\geq 0 \qquad (65)$$

$$=\frac{1}{2\pi j}\int_{-\Gamma_1}\Phi(s)e^{st}ds \quad \text{for } t<0$$

This is the definition used in reference 1.

A third definition is

$$\Phi_c{}^*(s)=\frac{1}{T}\sum_{n=-\infty}^{\infty}\Phi(s+jn\omega_s) \qquad (66)$$

and $\Phi_0(z)$ is obtained by substituting z for e^{sT} in $\Phi_c{}^*(s)$. This is the definition used in deriving equations 35 and 40. The series of equation 64 is convergent on the unit circle of the z-plane, or in the neighborhood of the imaginary axis on the s-plane (shown as region II in Fig. 6). In both reference 1 and in this paper, integrals involving $\Phi_0(z)$ and $\Phi_c(z)$ are evaluated along the unit circle.

It will be shown that the three definitions of $\Phi(s)$ are identical. In other words, when $\Phi_a(s)$, $\Phi_b(s)$, and $\Phi_c(s)$ are expressed in closed form, they are one and the same function.

The method follows closely that of Ragazzini and Zadeh,[4] and Jury,[7] by which the conventional z-transform is derived. Let $\Phi_a{}^*(s)$, $\Phi_b{}^*(s)$, and $\Phi_c{}^*(s)$ represent the respective sampled Laplace transforms. Let

$$\delta_{T^+}(t)=\sum_{n=0}^{\infty}\delta(t-nT) \qquad (67)$$

and

$$\delta_{T^-}(t)=\sum_{n=1}^{\infty}\delta(t+nT) \qquad (68)$$

The Laplace transform of equation 67 gives

$$\Delta_{T^+}(s)=\sum_{n=0}^{\infty}e^{-nsT} \qquad (69)$$

The series on the right-hand side of equation 69 is convergent for positive values of s. In the region of convergence, it can be expressed in closed form as

$$\Delta_{T^+}(s)=\frac{1}{1-e^{-sT}} \qquad (70)$$

Similarly, for negative values of s, the Laplace transform of equation 68 gives

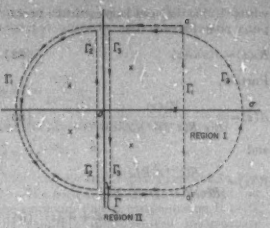


Fig. 6. Diagram showing pole locations, contours of integration, and regions of convergence

$$\Delta_{T^-}(s)=\sum_{n=1}^{\infty}e^{snT}=\frac{e^{sT}}{1-e^{sT}}=-\frac{1}{1-e^{-sT}} \qquad (71)$$

From equations 62, 63, and 69, the conventional z-transform can be written as

$$\Phi_a{}^*(s)=\mathcal{L}[\varphi_0(t)\delta_{T^+}(t)]$$

$$=\frac{1}{2\pi j}\int_{\Gamma_1}\Delta_{T^+}(s-s')\Phi(s')ds' \qquad (72)$$

where s' is the dummy variable of integration, and the path of integration Γ_1 is traversed by s'. Equation 72 is a convolution integral in the s-domain. For value of s to the right of aa', $s-s'$ is positive along the entire path of integration. Therefore writing $\Delta_{T^+}(s-s')$ in closed form according to equation 70 is allowable.

$$\Phi_a{}^*(s)=\frac{1}{2\pi j}\int_{\Gamma_1}\frac{\Phi(s')ds'}{1-e^{-(s-s')T}}$$

(for s in region I) (73)

Similarly, from equations 64, 65, 70, and 71

$$\Phi_b{}^*(s)=\mathcal{L}[\varphi_0(t)\delta_{T^+}(t)]+\mathcal{L}[\varphi_0(t)\delta_{T^-}(t)]$$

$$=\frac{1}{2\pi j}\int_{\Gamma_2}\frac{\Phi(s')ds'}{1-e^{-(s-s')T}}+$$

$$\frac{1}{2\pi j}\int_{\Gamma_3}\frac{\Phi(s')ds'}{1-e^{-(s-s')T}} \qquad (74)$$

Equation 74 holds for s in region II, i.e., the space in between Γ_2 and Γ_3. With s in region II, the conditions of convergence for both equations 70 and 71 are satisfied.

In evaluating the contour integrals, considering s as being in region I in equation 73 means simply disregarding the poles at $(s-s')T=j2\pi n$, where n is an arbitrary integer; and considering s as being in region II in equation 74 means also the same. Since $\Phi(s)$ does not have any pole in region II, Γ_2 and Γ_3 of equation 74 can be combined to give Γ_1. Therefore

$$\Phi_a{}^*(s)=\Phi_b{}^*(s) \qquad (75)$$

To prove that $\Phi_a{}^*(s)$ is also the same function, consider an integral with a large circle of infinite radius as the contour of integration Γ_∞

$$\frac{1}{2\pi j}\int_{\Gamma_\infty}\frac{\Phi(s')ds'}{1-e^{-(s-s')T}}=0 \qquad (76)$$

The integral vanishes since $\Phi(s')$ approaches zero at least as fast as $1/s^2$. The left-hand side of equation 76 can be considered as the sum of two integrals, J_1 and J_3. In evaluating J_1, the poles at $(s-s')T=j2\pi n$ are disregarded, and in evaluating J_3, the poles of $\Phi(s')$ are disregarded. Therefore

$$\Phi_a{}^*(s)=J_1=-J_3 \qquad (77)$$

Let $\omega_s=2\pi/T$. The residue at $s'=s+jn\omega_s$ is $-\Phi(s+jn\omega_s)/T$. Therefore

$$J_3=-\frac{1}{T}\sum_{n=-\infty}^{\infty}\Phi(s+jn\omega_s)=-\Phi_c{}^*(s) \qquad (78)$$

From equations 75, 77, and 78, the following equality is established

$$\Phi_a{}^*(s)=\Phi_b{}^*(s)=\Phi_c{}^*(s) \qquad (79)$$

This is the same as

$$\Phi_a(s)=\Phi_b(s)=\Phi_c(s)=\Phi(s) \qquad (80)$$

The foregoing proof justifies the use of standard z-transform tables for the calculation of $\Phi(s)$. While it is proved only for the case where $\Phi(s)$ does not have any pole on the $j\omega$ axis, the result is nevertheless applicable to cases in which $\Phi(s)$ has a finite number of poles on the $j\omega$ axis, in the limiting sense that these poles can be approximated to any degree of desired accuracy by poles not exactly on the $j\omega$ axis, as discussed in reference 1 and in this paper.

Appendix III. An Optimum System Designed for Absolute Zero Error After a Few Sampling Periods

Let $V_1(s)$ and $U(s)$ represent two polynomials of s^{-1} such that

$$G(s)=\frac{V_1(s)}{U(s)}$$

If $K(z)=W(z)V_1(z)$, where $W(z)$ is either a constant or a polynomial in z^{-1}, and $K(1)=1$, the system error due to a step input vanishes after n or m sampling periods, whichever is larger, where n and m are the highest powers of z^{-1} in $K(z)$ and $W(z)U(z)$ respectively. If $K(z)=W(z)V_1(z)$, $K(1)=1$, $K'(1)=0$, and $U(z)$ has at least a single root at unity, the system error due to a ramp input vanishes after n or m sampling periods, whichever is larger.

The foregoing result was first published by Kalman in a less general form.[8] The present version can be proved on the assumption that the plant transfer function $G(s)$ does not possess poles which are located $j2\pi/T$ apart on the s-plane. Without going into detail, the outline of the proof is as follows: The system error at sampling instants is $E(z)=R(z)[1-K(z)]$. Since $R(z)[1-K(z)]$ is a polynomial of z^{-1} with highest power $n-1$, the error vanishes at the nth subsequent sampling instant and thereafter. Similarly, the input $E_0(s)$ to the hold circuit is $R(z)W(z)U(z)$. It is either zero or a constant at the mth subsequent sampling instant and thereafter. Consequently, the input to the plant is either 0 or a constant after $t=mT$, and the plant output can be written as $r(t)$ plus a few transient terms. Since the system error is

zero at $t=nT$, $(n+1)T$, ..., etc., the transient terms must have zero coefficients. This proves the opening statements.

Consider the design problem of Appendix I.

$$V_1(z)=z^{-1}(1+2.34z^{-1})(1+0.16z^{-1})$$

Since there is no other restriction, and fastest settling is desired, $W(z)$ should be a polynomial of lowest order which satisfies $K(1)=1$, and $K'(1)=0$. Therefore

$$W(z)=0.732(1-0.647z^{-1})$$

$$1-K_0=(1-z^{-1})^2(1+1.268z^{-1}+0.177z^{-2})$$

$$=(1-z^{-1})^2(1+1.108z^{-1})(1+0.16z^{-1})$$

$$D(z)=\frac{0.732(1-0.647z^{-1})(1-0.368z^{-1})^2}{(1-z^{-1})(1+1.108z^{-1})(1+0.16z^{-1})}$$

References

1. Statistical Design Theory for Strictly Digital Sampled-Data Systems, **S. S. L. Chang.** *AIEE Transactions*, vol. 76, pt. I, 1957 (Jan. 1958 section), pp. 702–09.

2. Synthesis and Critical Study of Sampled-Data Control Systems, **Eliahu I. Jury.** *Ibid.* vol. 75, pt. II, July 1956, pp. 141–51.

3. The Design of Sampled-Data Feedback Systems, **Gladwyn V. Lago, John G. Truxal.** *Ibid.*, vol. 73, pt. II, Nov. 1954, pp. 247–53.

4. Control System Synthesis (book), **John G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955, p. 503.

5. Sampled-Data Processing Techniques for Feedback Control Systems, **Arthur R. Bergen, John R. Ragazzini.** *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 236–47.

6. The Analysis of Sampled-Data Systems, **J. R. Ragazzini, L. A. Zadeh.** *Ibid.*, vol. 71, pt. II, Nov. 1952, pp. 225–34.

7. Analysis and Synthesis of Sampled-Data Control Systems, **Eliahu I. Jury.** *Ibid.*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

8. *Discussion* by **R. E. Kalman** of reference 5, pp. 245–46.

9. Additions of z-Transformation Theory for Sampled-Data Systems, **Gladwyn V. Lago.** *AIEE Transactions*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 403–08.

Discussion

A. S. Robinson (Eclipse-Pioneer, Division of Bendix Aviation Corporation, Teterboro, N. J.): It appears important to note that the synthesis equations presented in this paper are not valid when either the poles or the zeros of $R(z)$ are outside the unit circle. The author's attention is directed to the work of Franklin.[1] In chapters 4 and 5 of his thesis Prof. Franklin presents solutions to the control system optimization problems considered in this paper. Prof. Franklin's results, partially rewritten for ease of comparison in the symbology used by Mr. Chang, give the optimum computer program for a control system with deterministic input and a minimum phase plant without transportation lag, as follows.

$$D(z)=\frac{A(z)}{1-A(z)G(z)} \tag{81}$$

where $A(z)$ is the open-loop equivalent computer program such that

$$K_0(z)=A(z)G(z) \tag{82}$$

Further,

$$A(z)=\frac{W_1(z)}{[R(z)R(z^{-1})]^+[\hat{G}G(z)]^+} \tag{83}$$

and

$$W(z)=\frac{R(z^{-1})\hat{G}RL(z)}{[R(z)R(z^{-1})]^-[\hat{G}G(z)]^-}=W_1(z)+W_2(z) \tag{84}$$

where the poles of $W_1(z)$ are all inside the unit circle and the poles of $W_2(z)$ are all outside. The expressions in brackets are defined by the example

$$\hat{G}G(z)=[\hat{G}G(z)]^-[\hat{G}G(z)]^+ \tag{85}$$

where the plus superscript indicates poles and zeros inside the unit circle, and the minus superscript indicates poles and zeros outside the unit circle.

With reference to Mr. Chang's results for this problem, substituting equation 82 in equation 9 yields

$$D(z)=\frac{A(z)G(z)}{G(z)[1-A(z)G(z)]} \tag{86}$$

which corresponds to equation 81. Substituting equation 82 in equation 8 gives

$$A(z)=\frac{Q_1(z)}{G(z)R(z)Y(z)} \tag{87}$$

But

$$Y(z)=\frac{[\hat{G}G(z)]^+}{G(z)} \tag{88}$$

and

$$Y(z^{-1})=\frac{[\hat{G}G(z)]^-}{G(z^{-1})} \tag{89}$$

Substitution of equation 89 in equation 6 gives

$$Q(z)=Q_1(z)+Q_0(z)=\frac{\hat{G}R(z)}{[\hat{G}G(z)]^-} \tag{90}$$

where the poles of $Q_1(z)$ are all inside the unit circle and the poles of $Q_0(z)$ are all outside. Mr. Chang's optimum computer program can therefore be written as

$$A(z)=\frac{Q_1(z)}{R(z)[\hat{G}G(z)]^+} \tag{91}$$

Prof. Franklin's results are not based on the assumption that the poles and zeros of $R(z)$ are all inside the unit circle. Under conditions where such an assumption is valid

$$[R(z)R(z^{-1})]^+=R(z) \tag{92}$$

and

$$[R(z)R(z^{-1})]^-=R(z^{-1}) \tag{93}$$

If, in addition, the desired response $L(s)$ is unity, equation 84 simplifies to

$$W(z)=\frac{\hat{G}R(z)}{[\hat{G}G(z)]^-} \tag{94}$$

which is identically equal to $Q(z)$ of equation 90. Further, equation 83 simplifies to

$$A(z)=\frac{W_1(z)}{R(z)[\hat{G}G(z)]^+} \tag{95}$$

which is identically equal to equation 91.

Comparable comparisons can be made for the equations governing the synthesis of systems with random inputs.

The writer has shown[2] that results of this type are actually applicable only to the synthesis of data-limited sampled-data systems. A system is said to be data limited when the system sampling period T is set by some constraint not under the control of the computer designer, such as the speed of rotation of a radar antenna that is deriving the input data. When a digital computer is used to actuate a continuous process the system sampling period is limited by the computer speed rather than by the data rate. Systems of this type are computer limited, not data limited. Each step in the digital computer program is performed in sequence, and requires a finite amount of computing and data-processing time. If a number of channels are to be actuated, each channel requires a computing time that is a function of the number of terms in the computer program for that channel. The over-all system sampling period T is therefore a function of the total number of program terms the computer is required to process for all channels. In general, any system that utilizes a digital computer in real time is a computer-limited sampled-data system, unless the available input data rate is slower than the speed with which computer solutions can be obtained.

Techniques required for the synthesis of computer-limited sampled-data control systems are summarized in a recent paper.[3]

In general, it is found that optimum computer-limited computer programs contain less terms and therefore require shorter sampling periods than programs indicated by data-limited theory.

References

1. The Optimum Synthesis of Sampled-Data Systems, **G. Franklin.** *D. Eng. Thesis*, Columbia University, New York, N. Y., May 1955.

2. The Optimum Synthesis of Computer Limited Sampled Data Systems, **Arthur S. Robinson.** *Ibid.*, May 1957.

3. The Synthesis of Computer-Limited Sampled-Data Control Systems, **Arthur S. Robinson.** "Proceedings of the Computers in Control Systems Conference," *AIEE Special Publication T101*, May 1958, pp. 77–87.

Paul M. DeRusso and **Frederick W. Nesline, Jr.** (Massachusetts Institute of Technology, Cambridge, Mass.): Prof. Chang is to be commended for the excellent design philosophy in this paper. The significant difference between this work and that of others[1,2] who have considered the same problem is that Prof. Chang has considered the effects of physical constraints on linear digital system design. This idea was introduced to the control field for continuous systems by Newton[3,4] and certainly should be applied to sampled-data systems. No realistic design technique can ignore saturation as a performance limitation.

The purpose of this discussion is to present some results of work performed on the same problem at Massachusetts Institute of Technology. By way of clarification, it should be noted that in the paper the

correlation function is defined by using a time average whereas the mathematical definition of the correlation function requires an ensemble average.[5-7] Ensemble averaging and time averaging are equivalent only when the functions are stationary and ergodic. Such is not the case here, however, even though the input to the system is a stationary stochastic signal, since sampling makes the ensemble average of the squared error a periodically time-varying function. A time average of the squared error eliminates the time dependence. Therefore, the results of this paper are equivalent to what one obtains if the ensemble average of the squared error is time-averaged over one sampling interval.[8] In other words, what is really being minimized is the time average of the mean-square error. However, in this case, minimization of the instantaneous mean-square error and minimization of the time average of the mean-square error result in the same compensation, since neither the input statistics nor the system parameters are time-varying.

In regard to the integral-square or time-averaged mean-square error which results from the design procedures of this paper, the evaluation of the integrals of equations 35 and 40 can be difficult in their present form. However, if the transformation

$$Z=\frac{1+W}{1-W}$$

is introduced, the integral tables[9] which have been developed for the analogous problem with continuous systems can be utilized,[8] resulting in a considerable saving of effort.

Finally, this paper has considered strictly digital compensation of error-sampled continuous systems. Less error results when an error-sampled system is compensated by a combination of digital and continuous filtering. In other words, use of the optimum "generalized hold" can result in a still further reduction of the mean-square error. The extent of this reduction, of course, depends upon a number of factors such as the fixed elements, sampling interval, input spectrum, noise, desired output, etc.

Methods have been developed for the determination of the combination of digital and continuous compensation which minimizes the time-averaged mean-square error.[8] As a simple example of the results of this type of design, consider the case where noise and saturation are neglected, and

$$L(s)=1$$

$$\Phi_{rr}(s)=\frac{2\nu\beta^2}{4\nu^2-s^2}$$

and the plant is a time delay

$$G(s)=e^{-sT_1}$$

(The symbols are defined according to Fig. 2.) If a normal hold circuit is specified and the optimum digital compensation is used, the resulting time-averaged mean-square error is

$$e^2(t)_{\text{dig}}=\beta^2\left\{1-e^{-\nu T_1}\left[\frac{1-e^{-2\nu T}}{2\nu T}\right]^2\right\}$$

On the other hand, if the hold system is unspecified and the optimum digital plus continuous compensation are used, the resulting time-averaged mean-square error is

$$e^2(t)_{\text{dig and cont}}=\beta^2\left\{1-e^{-\nu T_1}\left[\frac{1-e^{-\nu T}}{4\nu T}\right]^2\right\}$$

The optimum continuous compensation is a simple exponential decay $e^{-2\nu t}$ and its use in this case could result in better than an 8% reduction of the time-averaged mean-square error from that which results when the normal hold is used. The exact value depends upon νT and T_1, and has a peak for this case at approximately $\nu T=4$. Thus, this type of compensation is most useful for the large values of T that may be necessary when the digital computer is time-shared among many processes.

References

1. The Optimum Synthesis of Sampled-Data Systems, J. R. Ragazzini. *Technical Report T-6/B*, Columbia University, New York, N. Y., 1955.

2. Random-Signal Analysis of Linear and Nonlinear Sampled-Data Systems, N. W. Trembath. *Report No. 108*, Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., 1957.

3. Compensation of Feedback Control Systems Subject to Saturation, G. C. Newton, Jr. *Journal*, Franklin Institute, Philadelphia, Pa. 1952, pp. 281–96, 391–413.

4. Analytical Design of Linear Feedback Controls (book), G. C. Newton, Jr., L. A. Gould, J. F. Kaiser. John Wiley & Sons, Inc., New York, N. Y., 1957.

5. An Optimization Theory for Time-Varying Linear Systems with Nonstationary Statistical Inputs, R. C. Booton, Jr. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Aug. 1952, pp. 977–81.

6. Newton, Gould and Kaiser, *op. cit.*, pp. 80–86.

7. Random Processes in Automatic Control (book), J. H. Laning, Jr., R. H. Battin. McGraw-Hill Book Company, Inc., New York, N. Y., 1956, chap. 3.

8. Ultimate Performance Limitations in Digitally Controlled Systems, P. M. DeRusso. *S. E. Thesis*, Massachusetts Institute of Technology, 1958.

9. Theory of Servomechanisms (book), H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, Inc., 1947, pp. 369–70.

R. E. Kalman (International Business Machines Corporation, Ossining, N. Y.): Prof. Chang's lucid and timely paper is an important contribution to the growing literature of sampled-data systems. As it often happens in an actively developing field, several other workers have studied simultaneously and independently the problems solved by the author. For instance, Franklin, in a doctorate dissertation, has formulated Wiener-like techniques for sampled-data systems.[1] This discusser (with Bertram) recently extended[2] an earlier remark[3] concerning sampled-data systems with deadbeat response into a complete design theory which is rigorous and more general than the author's brief outline in Appendix III. In addition, this discusser (with Koepcke) has recently applied the method of dynamic programing, originated by Bellman,[4] to the solution of problems of the Wiener type. An exposition of this method, as applied to the design of sampled-data systems which have optimum response (in the sense of a quadratic performance index) to certain classes of input signals, will appear shortly.[5] The following brief remarks are intended to compare some results of the discusser's method with some of the results of the author.

Suppose the system to be optimized has the open-loop transfer function, in the author's notation (see Fig. 1),

$$G(s)=\frac{1-e^{-\pi s/2}}{s}\times\frac{1}{s(s^2+0.6s+1.09)}$$

Let $c(kT)$ be the output and $e(kT)$ be the error at the sampling instants, with $T=\pi/4$. Consider the following performance indexes:

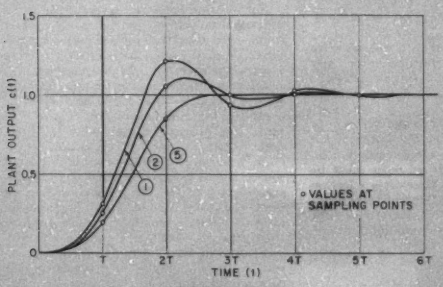


Fig. 7. Unit step response

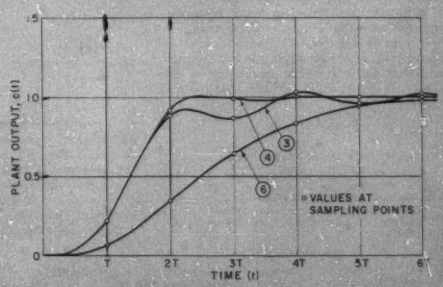


Fig. 8. Unit step response

$$J_1=\sum_{k=0}^{\infty} e^2(kT)$$

$$J_2=\sum_{k=0}^{\infty} 4^k e^2(kT)$$

$$J_3=\sum_{k=0}^{\infty} e^2(kT)+0.5\left(\frac{de(kT)}{dt}\right)^2$$

$$J_4=\sum_{k=0}^{\infty} 4^k\left[e^2(kT)+0.5\left(\frac{de(kT)}{dt}\right)^2\right]$$

$$J_5=N$$

where N is an integer such that $e^2(t)=0$ if $t \geq NT$ (deadbeat response),

$$J_6=\sum_{k=0}^{\infty} e^2(kT)+\left(\frac{de(kT)}{dt}\right)^2+\left(\frac{d^2e(kT)}{dt^2}\right)^2$$

Similarly, one can define performance indexes in which the summations are replaced by integration.[3]

The system is optimized by choosing $D(z)$ (see Fig. 1) in such a way that the various performance indexes become a minimum. The actual calculations to obtain the coefficients specifying $D(z)$ are carried out by means of an iterative procedure (see reference 5 for details). The response to step inputs of systems designed in such a way is shown in Figs. 7 and 8. In Fig. 7, it is seen that the response becomes less oscillatory and better damped if the squared errors are weighted in the performance index. Still better damped is the response of the system designed to deadbeat (J_5).

Fig. 8 indicates an attempt to design a system whose step response, in an intuitive sense, is better than that shown in Fig. 7. This is achieved by hunting for a different performance index. It is seen that if the terms in performance index J_3 are weighted by a factor which increases exponentially (J_4), then we obtain the very fast and very well damped response indicated by 4 in Fig. 8.

In conclusion, it should be noted that dynamic programing provides an alternate tool for solving problems of the Wiener type. Not only does this method allow the use of a very wide variety of performance indexes (which include all of those customarily used so far in the literature and many more which have not been used yet) but the numerical computation of the optimal design is greatly facilitated. For these reasons, it is the discusser's view that dynamic programing may prove to be a better tool for the solution of Wiener-type optimization problems than current methods which are based on the concepts of a transfer function and spectral factorization.

References

1. See reference 1 of Dr. Robinson's discussion.

2. General Synthesis Procedures for Computer Control of Single and Multiloop Linear Systems, R. E. Kalman, J. E. Bertram, "Proceedings of the Computers in Control Systems Conference," *AIEE Special Publication T101*, May 1958.

3. See reference 5 of the paper.

4. An Introduction to Dynamic Programming (book), R. E. Bellman. Princeton University Press, Princeton, N. J., 1957.

5. Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes, R. E. Kalman, R. W. Koepcke. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., 1958 (to be published).

S. S. L. Chang: The author wishes to thank the discussers for the various interesting and pertinent points raised, and for calling his attention to Dr. Franklin's excellent thesis.[1] Perhaps the best way of acknowledging Dr. Franklin's work is to compare objectively his work and the author's:

1. The differences are mostly in the methods of approach and not so much in the results. Franklin's error integrals are derived entirely by convolution in the time domain. The function $A(z)$ (referred to in Dr. Robinson's discussion) is expressed in a time series, the optimization is carried out term by term, and the result is expressed as an infinite sum though obviously it can be converted to closed form. The author's time-domain manipulations end as soon as the basic transmission properties of equations 24 and 31 are derived, the optimization is carried out by considering pole-zero locations on the z-plane, and the result is in closed form without further ado. While Franklin's time-domain approach is an exact parallel to Wiener's original work, the author's z-domain approach may have the dubious advantages of greater flexibility and less mathematical gymnastics.

2. Franklin's result is equally applicable to systems with transportation lag and nonminimum phase, but not if the plant is unstable. The author's result is applicable under all the aforementioned contingencies up to the point of deriving $K_0(z)$. However, for an unstable plant, a slightly modified procedure for synthesizing the final system from $K_0(z)$ is called for.

Before answering the various questions raised by the discussers, it is perhaps desirable to clarify two points in the paper itself: 1. treatment of unstable plants, and 2. treatment of deterministic inputs.

Unstable Plant

Bertram[2] has shown that for a series-compensated system with unity feedback, and pole of $G(z)$ outside the unit circle must appear as a zero of $1-K(z)$. This restriction can be removed by either z-dependent feedback or shunt compensation. For instance, the digital processing unit can be expressed as

$$E_0(z)=D_1(z)E(z)+D_2(z)C(z) \tag{96}$$

Knowing $K(z)$, $D_1(z)$ and $D_2(z)$ are not completely determined. However, for any $K(z)$, a pair of $D_1(z)$ and $D_2(z)$ can always be found which do not cancel the poles of $G(z)$ directly.

The next consideration is the selection of an unknown function in the optimization process. The only allowable restriction on this function is that all its poles must be inside the unit circle. Franklin's selection of the equivalent series function $A(z)$ automatically satisfies the restrictions for transportation lag and nonminimum phase. Since

$$K(z)=A(z)G(z)$$

any zero of $G(z)$ outside the unit circle appears in $K(z)$. However, additional provision must be made for poles of $G(z)$ outside the unit circle. Similarly, if $K(z)$ is selected as the unknown function, poles of $G(z)$ outside the unit circle do not present a problem. However, additional provision must be made for the zeros. To derive a synthesis procedure good for all situations, it is a choice between two evils. The latter choice is made in this paper with all the unavoidable discussion on transportation lag and nonminimum phase.

Next, the applicability of equations 24 and 31 must be considered. In deriving these equations, it was stated that $G(z)$, $G_0(z)$ and $G_2(z)$ represent stable network functions. In all appearances, this condition is violated by an unstable plant. However, closer examination shows that it is not. Let a_i, with $i=1, 2 \ldots, n$, represent all the poles of the plant transfer function $G(z)$ with positive real components. Then

$$G(z)=\frac{G_2(z)}{\prod_{i=1}^{n}(z-a_i)} \tag{97}$$

$$G^*(z)=\frac{G_2^*(z)}{\prod_{i=1}^{n}(1-e^{a_iT}z^{-1})} \tag{98}$$

where all the poles of $G_2(z)$ and $G_2^*(z)$ have negative real components. Equation 41 can be written as

$$C(z)=R_1(z)\times\frac{K^*(z)}{G_2^*(z)}\times G_0(z) \tag{99}$$

where

$$G_0(z)=\left(\prod_{i=1}^{n}\frac{1-e^{a_iT}z^{-1}}{z-a_i}\right)G_2(z) \tag{100}$$

In equation 100, each factor in the product has an impulse response function $f_i(t)$:

$$f_i(t)=e^{a_it} \qquad 0<t<T$$

$$f_i(t)=0 \qquad t<0, t>T$$

The impulse response function $g_0(t)$ is obtainable by convolution of all the $f_i(t)$'s and $g_2(t)$. Obviously, $g_0(t)$ vanishes as t approaches infinity. With $K(z)$ defined as it is in equation 51, all the poles of $K(z)/G_2(z)$ are inside the unit circle.

The function $K(z)/G_2(z)$ is the $H(z)$ of Fig. 5(A), and $G_0(z)$ is the $G(z)$ of Fig. 5(A) as well as the $G_0(z)$ and $G_2(z)$ of Fig. 5(B).

Deterministic Inputs

For a sampled-data system, the term deterministic inputs has a much stricter implication that that for a continuous system. With the obvious exception of step inputs, the starting time must be fixed in relation to the sampling instant. Otherwise, the input can at best be represented as a random combination of deterministic inputs, but cannot be treated as a single deterministic input.

As it is inconceivable that an external source is synchronized to the sampling of a control system, the foregoing consideration limits the application of the optimum design procedure for deterministic inputs given in this paper, as well as Franklin's results mentioned in Dr. Robinson's discussion, to test functions only. An unsynchronized deterministic input can be treated in exactly the same way as a random input (case 2) with $\Phi_{rr}(s)=R(s)R(-s)$.

In reply to Dr. Robinson's comment about strictly deterministic inputs with poles and zeros of $R(z)$ outside the unit circle, the discusser's attention is called to equations 54 and 55. In order that equation 55 be separable into two types of terms according to the location of poles inside and outside the unit circle respectively, and with the unknown functions $F_0(z)$ and $X(z)$ in the two separate parts, it is obvious that equation 54 must be divided by the modified form of $R(z^{-1})$ instead of $R(z^{-1})$ itself. There is then no difference between Franklin's and the authors results. However, since none of the standard test functions, i.e., step, ramp, or acceleration, has poles or zeros outside the unit circle, the point is: why bother?

Dr. Robinson's concept of trading the number of operations for sampling rate in computer-limited systems deserves much commendation. However, the author must admit that he has some reservation about the general workability of the synthesis procedure proposed in Dr. Robinson's paper.[3] The simultaneous equations which are obtained by differentiating the mean-square error with respect to the N unknown coefficients appear to be of high order with the unknown coefficients intermingled in each term. It is doubtful if these equations can be solved except in trivial cases. The use of a digital computer for solving these equations appears to be a simple way out; however, without an analytical program, an N-dimensional combing operation may easily take a computer anywhere from a few minutes for trival systems to billions of years for moderately complicated systems with N equal to ten or even less. Some clarification by Dr. Robinson on how the simultaneous equations can best be solved is perhaps worthwhile.

Incidentally, the method described in the paper can be readily extended to synthesize approximately optimum sampled-data systems with reduced number of computer operations. While the main purpose of the procedure is to bring out the ultimate capabilities of the controlled plant, some compromise is made for controller economy or feasibility by trading a small measure of performance for reduced number of computing operations. The steps are as follows:

1. The least-mean-square error $\bar{\epsilon}^2$ or the least cost $\bar{\epsilon}^2+\eta\bar{v}^2$ is calculated as a function of T without going through the complete design process. It is plotted and a compromise value of T is selected.
2. $K_0(z)$ and $D(z)$ are calculated.
3. $D(z)$ usually embodies considerable cancellation of $G(z)$.

As this is known as a rather wasteful way of compensating a system, fewer poles and zeros are used in compromise program $D_1(z)$ to do approximately the same job. The requirement that the new closed-loop function $K_1(z)$ duplicates or nearly duplicates the significant poles and zeros of $K_0(z)$ determines $D_1(z)$. In general, the poles or zeros of $K_0(z)$ clustered around or near the origin are insignificant.[4]

One crucial consideration for approximating the closed-loop poles and zeros is, "how close is close?" Knowing the allowable increase in mean-square error, what are the tolerance regions of pole-zero locations? An equation relating the displacements and increase in mean-square error is desirable.

From equations 13, 40, 50, 55, 56, 58, and 61 it can be readily shown that

$$(\bar{\epsilon}^2+\eta\bar{v}^2)_{\text{optimum}}=\frac{1}{2\pi j}\oint\left[\mathfrak{L}L\Phi_{rr}(z)-\frac{Q_f(z)Q_f(z^{-1})}{T}\right]\frac{dz}{z} \quad (101)$$

$$\delta(\bar{\epsilon}^2+\eta\bar{v}^2)=\frac{1}{2\pi jT}\oint\Phi_{rr1}(z)Y(z)Y(z^{-1})\times\delta F(z)\delta F(z^{-1})\frac{dz}{z} \quad (102)$$

Equation 102 is an exact equation, and $\delta F(z)=F_1(z)-F_0(z)$, where $z^{-m'}P(z)F_1(z)$ is the closed-loop function of the economized system. When the poles and zeros of F_1 and F_0 are near enough, the following approximate form may be used

$$\delta F(z)\cong F_0(z)\delta\ln F(z) \quad (103)$$

$$F_0(z)=\frac{\prod_i^{\gamma}(1-c_iz^{-1})}{\prod_i^{\gamma}(1-b_iz^{-1})} \quad (104)$$

$$\delta\ln F(z)=\sum_i\frac{\delta b_i}{z-b_i}-\sum_j\frac{\delta c_j}{z-c_j} \quad (105)$$

where b_i's and c_i's are poles and zeros of $F_0(z)$, and δb_i, δc_j represent the respective displacements. Equation 102 can be written as

$$\delta(\bar{\epsilon}^2+\eta\bar{v}^2)=\frac{1}{2\pi jT}\oint Q_f(z)Q_f(z^{-1})|\delta\ln F(z)|^2\frac{dz}{z} \quad (106)$$

The significance of equations 101 and 106 are as follows: Once an optimum system is selected, the constant η is also known, and $\bar{\epsilon}^2+\eta\bar{v}^2$ can be viewed as a total cost. If the economized system has the same $\bar{\epsilon}^2$, then $\delta(\bar{\epsilon}^2+\eta\bar{v}^2)$ is the same as $\delta\bar{\epsilon}^2$; otherwise, $\delta(\bar{\epsilon}^2+\eta\bar{v}^2)$ is a more complete measure. Equation 101 shows that the integral

$$J_q=\frac{1}{2\pi j}\oint\frac{Q_f(z)Q_f(z^{-1})}{T}\frac{dz}{z} \quad (107)$$

represents the optimum effort of the system. For a system of good performance, J_q must be only slightly smaller than the mean-square value of desired output $\bar{C}_i^2$.

The potential analog of equations 105 and 106 are evident. Each displacement δb_i and δc_j can be considered as a dipole moment. The magnitude of the complex potential introduced by these dipole moments at the unit circle must be kept as low as possible, especially at the points where the magnitude of $Q_f(z)$ is large. A quick estimate of the increased cost is:

$$\frac{\delta(\bar{\epsilon}^2+\eta\bar{v}^2)}{\bar{C}_i^2}<\frac{\delta(\bar{\epsilon}^2+\eta\bar{v}^2)}{J_q}<\text{maximum value of }|\delta\ln F(z)|^2\text{ on the unit circle} \quad (108)$$

Inequality 108 can be improved in case $\Phi_{rr1}(z)$ has poles on or near the unit circle.[5]

From equation 104 it can be shown readily that if a pole b_i or a zero c_i of $F_0(z)$ in the vicinity of the origin does not appear in $F_1(z)$, the displacements δc_i and δb_i can be taken as $-c_i$ and $-b_i$ respectively.

Once the tolerance regions are determined, the allowed poles and zeros of $D(z)$ are selected to give a $K(z)$ or $F_1(z)$ meeting these requirements. A zero of $D(z)$ is also a zero of $K(z)$ if it is not cancelled out. The poles are more difficult to manage. However, well-developed techniques for continuous systems[6] can be used.

Dr. Robinson's technique works best in case of severe computer limitation; however, if the main concern is to avoid overdesign, the author's approximation technique may have an "edge" in its simplicity.

From the comments of Mr. DeRusso and Dr. Nesline, one can easily surmise that Mr. DeRusso has written an interesting and valuable thesis. Their comment about the definition of the correlation function is very pertinent. One may add that though the author's definition is not in complete accord with the conventional usage of the word, it is nevertheless self-consistent and is in accord with the rest of the text.

While the normal hold circuit is not always desirable, as it represents a time lag of about half a sampling period, it cannot be discarded without added cost in hardware as well as in increased plant saturation. The plant always saturates worse with a sharper pulse than with a flat one.

Dr. Kalman's references to his recent work are most welcome. The deadbeat systems are given only a brief treatment because it is a side issue in this paper, no matter what its inherent worth may be. The author would like to add that any reader interested further on the subject of deadbeat systems can find a good source in the Kalman-Bertram work.[7]

The transient performance curves for least-square error of Fig. 3 and of Dr. Kalman's discussion are calculated without considering power constraint. As it turns out, equation 13 has the effect of not only controlling $v(t)$ but also giving a better damped and faster settling system.

References

1. See reference 1 of Dr. Robinson's discussion.
2. Factors in the Design of Digital Controllers for Sampled-Data Feedback Systems, John E. Bertram. *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 151–59.
3. See references 2 and 3 of Dr. Robinson's discussion.
4. Correlation Between Root-Locus and Transient Response of Sampled-Data Control Systems, Eliahu I. Jury. *AIEE Transactions*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 427–35.
5. Optimum Design Techniques for Feedback Control Systems (book), S. S. L. Chang. McGraw-Hill Book Publishing Company, Inc., New York, N. Y. "Tolerance Limits of Optimum Systems," chap. 8 (to be published).
6. See reference 4 of the paper, pp. 306–15.
7. See reference 2 of Dr. Kalman's discussion.

Approximate Analysis of the Drag Cup A-C Tachometer

L. A. KNOX
NONMEMBER AIEE

Synopsis: Temperature coefficient, output voltage, and speed linearity error of output voltage are indicated in terms of relative changes in tachometer cup impedances when the output winding is essentially unloaded. By replacing the drag cup of a precision drag-cup a-c tachometer with one having a higher resistivity and a lower temperature coefficient of resistance, the following improved characteristics were observed: 1. The output voltage temperature coefficient was reduced to −0.06%/1 C (degree centigrade) 1/4 its former value. 2. Up to synchronous speed, the output voltage speed linearity error was reduced to ±0.05% of its synchronous speed output, 1/12 its former error. And 3 output volts per 1,000 rpm was reduced to 0.8 volt, 2/7 its former value. A simplified basic tachometer performance equation is derived, and curves are presented to show the required change in drag cup parameters to realize a specified performance. Frequency compensation is achieved by an *R-L-C* network in the tachometer excitation circuit.

PRECISION long-range navigation for high-speed aircraft generally requires the integration of velocity; e.g., ground speed. A simplified schematic of a velocity integrator incorporating a motor tachometer is shown in Fig. 1. The integrator circuit consists of a signal amplifier driving a motor with a shaft coupled to the drag-cup a-c tachometer whose output voltage provides negative feedback to the signal amplifier. Provided that the amplifier gain is high, the tachometer will operate at a speed such that its output voltage is essentially equal (and opposite in polarity) to the input signal voltage. The number of revolutions of the shaft represents the integral of the input voltage.

Commercially available precision motor tachometers are inadequate for use as integrators in some high-speed aircraft applications where a wide range of speeds, temperatures, and supply frequencies must be accommodated. If the integration errors must be less than 0.1%, the range of operating speeds is severely limited (to 1/3 synchronous speed, approximately) and the permissible variation in temperature is extremely small (±1/4 C). A survey of the literature indicates that, to minimize these limitations, the drag cup material should have a high resistivity and a low temperature coefficient.[1,2] However, the degree of improvement to be expected from a given change in drag-cup material is not immediately apparent from these complex equations.

The tachometer performance equation parameters are presented here as normalized curves and simple ratios to increase their utility. The results of this analysis show that even "precision" motor tachometers may be improved, possibly by as much as a factor of ten.

An exploded view of a drag cup a-c tachometer is shown in Fig. 2. Fig. 3 shows a functional representation of a simple 2-pole tachometer.

Nomenclature

E_1 = volts excitation = $|E_1|_{max} \cos(\omega_1 t)$
E_0 = volts output
$(i_1)_0$ = amperes excitation current (cup removed)
i_2 = amperes output winding load current
i_R = effective rotor current associated with ϕ_{RO}
K_2 = constant
L_1 = henrys inductance of excitation winding (cup removed)
L_C = henrys effective rotor inductance associated with current i_R producing flux ϕ_{RO}
M = the speed term in equation 1, and is a function of both speed V and cup parameter $(\omega_1 L_C/R_C)$. The latter causing the speed nonlinearity of output voltage; see equations 18 and 19
N_1 = number of turns of excitation winding
N_2 = number of turns of output winding
$R_1 = (R_{1F} + R_{1C})$
R_{1C} = ohms referred to excitation winding (cup removed) based on copper resistance
R_{2C} = ohms referred to output winding (cup removed) based on copper resistance
R_{1F} = ohms referred to excitation winding (cup removed) based on iron loss
R_{2F} = ohms referred to output winding (cup removed) based on iron loss
R_g = reluctance of air gap
R_C = ohms effective rotor resistance associated with current i_R producing flux ϕ_{RO}
V = operating speed/synchronous speed
Z_O = ohms output load impedance
ϕ_1 = flux vector excitation winding with drag cup removed
ϕ_2 = flux vector from output winding load current
ϕ_{RO} = flux vector of rotor along output winding axis
α = degrees phase angle associated with factor M
$\frac{\epsilon_L}{3}$ = ±approximate speed error from median linear output line, e.g. ±rpm error tolerance/maximum rpm; refer to equation 2

The basic tachometer equation 1 (derived in the Appendix) is as follows:

$$E_0 \simeq \left(\frac{-4\pi^2 10^{-18}}{K_2}\right)\left[\frac{N_{2O1}}{N_1 R_g R_c}\right] \times \left(1 - j\frac{\frac{M}{R_1}}{\omega_1 L_1}\right) |E_1|_{max} \cos(\omega_1 t + \alpha) \quad (1)$$

An additional simplification is achieved by interpreting this equation in the form:

$$\frac{\text{Required performance of modified tachometer}}{\text{Actual performance of existing tachometer}}$$

Fig. 4 shows a typical plot of output voltage E_0 versus excitation frequency. The output voltage magnitude is directly proportional to the excitation frequency only at low frequencies. At higher frequencies, the output voltage reaches a maxima and then decreases. The phenomena may be caused by winding-distributed capacitance neglected in equation 1. Therefore, to avoid unnecessary complication of this equation, the variation of output voltage with variations of excitation frequency will be determined on an empirical basis.

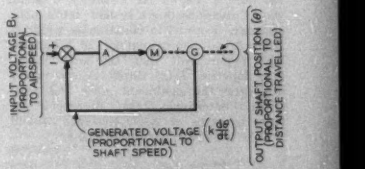


Fig. 1. Velocity integrator, amplifier-motor-generator type

Paper 58-85, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted April 12, 1957; made available for printing January 2, 1958.

L. A. KNOX is with the International Business Machines Corporation, Owego, N. Y.

The author gratefully acknowledges the co-operation of the Kearfott Company both in consultations and in supplying motor tachometers with aluminum silicon bronze drag cups. The measured performance of these units was used to check the theoretical equations presented in this paper. Mr. Bernard Tostanoski, by means of an IBM Digital Computer, evaluated factors M and α used in plotting the curves. His assistance as well as the efforts of Mr. Evan Evans in the preparation of this manuscript are appreciated.

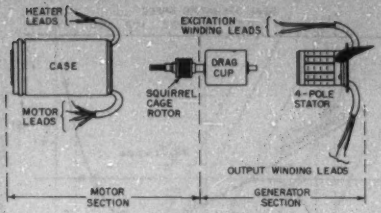


Fig. 2. Motor tachometer, exploded view

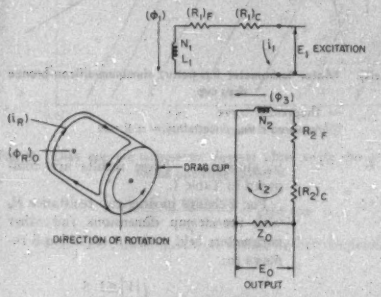


Fig. 3. Drag cup a-c tachometer

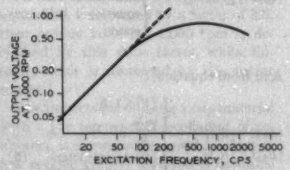


Fig. 4 (left). Output voltage E_0 versus excitation frequency

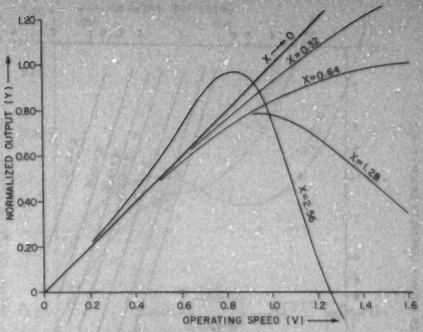


Fig. 5. Drag-cup a-c tachometer normalized output, in-phase component Y versus speed V

$Y_0 = V$ = normalized ideal linear output

for $\frac{\omega_1 L_c}{R_c} \to 0$

$$M/\underline{\alpha} = \left\{ \frac{(1+V)}{\left[1+j(1+V)\frac{\omega_1 L_c}{R_c}\right]} + \frac{-(1-V)}{\left[1+j(1-V)\frac{\omega_1 L_c}{R_c}\right]} \right\}$$

$$Y = \left[\frac{M \cos(\alpha - \alpha_0)}{\left.\frac{dM}{dV}\right|_{V\to 0}} \right]$$

$$X = \frac{\omega_1 L_c}{R_c} = \text{drag cup} \left[\frac{\text{reactance}}{\text{resistance}}\right]$$

α_0 = phase angle at $(V \to 0)$

$\left.\frac{dM}{dV}\right|_{V\to 0}$ = derivative as $(V \to 0)$

$V = \frac{\text{operating, rpm}}{\text{synchronous, rpm}}$

Application

The relative change in performance of a tachometer resulting from relative changes in its parameters may be determined from equation 1, provided that quantities M and α may be evaluated for the particular operating speed V and drag cup parameter

$$\left(\frac{\omega_1 L_c}{R_c}\right)$$

under consideration. These terms as defined by equations 18 and 19 have been evaluated by means of an IBM (International Business Machines Corporation) digital computer and plotted in Fig. 5.

Fig. 5 shows normalized output (in-phase component) versus operating speed V for various values of drag cup parameter

$$\left(\frac{\omega_1 L_c}{R_c}\right)$$

These curves were all normalized to an initial slope of unity at $(V \to 0)$. Lower values for the parameter

$$\left(\frac{\omega_1 L_c}{R_c}\right)$$

result in improved speed linearity; i.e., a closer conformity with the ideal linear output curve for

$$\left(\frac{\omega_1 L_c}{R_c} \to 0\right)$$

Fig. 6 shows the error or difference $(Y - Y_0)$ between the in-phase component of output voltage and the ideal linear output for normalized curves as shown in Fig. 5. The linearity error ϵ_L may be expressed as

$$\epsilon_L = \left[\frac{Y - Y_0}{Y_0}\right] \qquad (2)$$

From the curve in Fig. $5(\omega_1 L_c/R_c) = 0.32$

for $V = 1.2$ and $\frac{\omega_1 L_c}{R_c} = 0.32$ $\Big]$ $Y = 1.067$, $Y_0 = 1.200$, $(Y - Y_0) = -0.133$, $\epsilon_L = -0.111$ or -11.1%

This result is presented more accurately in Fig. 6 as a family of curves, each representing a particular value of drag cup parameter

$$\left(\frac{\omega_1 L_c}{R_c}\right)$$

A study of these curves indicates that the following approximation may be made:

$$\frac{(Y-Y_0)2}{(Y-Y_0)1} \simeq \left[\frac{\left(\frac{\omega_1 L_c}{R_c}\right)_2}{\left(\frac{\omega_1 L_c}{R_c}\right)_1}\right]^3 \quad \text{for} \quad \begin{cases} \frac{\omega_1 L_c}{R_c} \le 0.16 \\ |V| \le 1.5 \end{cases} \qquad (3)$$

Fig. 7 shows linearity error rpm versus tachometer ideal rpm as measured for a

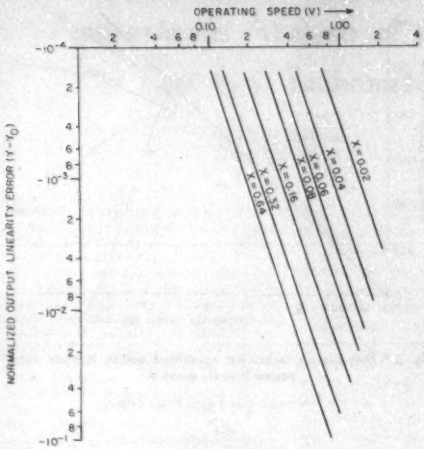


Fig. 6. Drag-cup a-c tachometer normalized output error, $Y-Y_0$ versus speed V

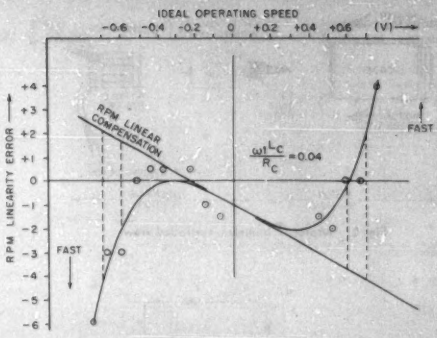


Fig. 7 (left). Motor tachometer integrator, aluminum-silicon-bronze drag cup

— Theoretical curve

⊙Measured data, uncertainty = ±2 rpm

$$\epsilon_L = \left[\frac{Y-Y_0}{Y_0}\right] = \left[\frac{Y-Y_0}{V}\right]$$

ϵ_L = output linearity total deviation

Y = normalized output in phase component

Y_0 = normalized ideal linear output

$$X = \frac{\omega_1 L_c}{R_c} = \text{drag cup}\left[\frac{\text{reactance}}{\text{resistance}}\right]$$

$$V = \frac{\text{operating rpm}}{\text{synchronous rpm}}$$

$\frac{\epsilon_L}{3}$ = ±linearity error from median linear output line

motor tachometer integrator with an aluminum silicon bronze drag cup. The error rpm intercept indicates improper compensation for zero input signal. A straight line drawn tangent to the error curve at zero rpm represents the speed compensation rpm error intentionally introduced. The rpm difference between these curves divided by the tachometer ideal rpm is the linearity error ϵ_L, without speed compensation. This is shown by equation 2. From Fig. 6, with the normalized error $(Y-Y_0)=(\epsilon_L V)$ evaluated and with both V and $Y-Y_0$ known, the drag cup parameter

$$\left(\frac{\omega_1 L_c}{R_c}\right)$$

may be read from the curves.

Similar measurements were made using an identical motor tachometer except that the drag cup material was changed to aluminum. These results are compared in Table I.

For a change in drag cup resistance R_c with the air gap dimensions and other parameters held constant, equation 3 reduces to:

$$\frac{(\epsilon_L)2}{(\epsilon_L)1} \simeq \left[\frac{(R_c)1}{(R_c)2}\right]^3 \text{ for } \begin{cases} |V| \le 1.5 \\ \dfrac{\omega_1 L_c}{R_c} \le 0.16 \\ \text{other parameters of equation 1 held constant} \end{cases} \quad (4)$$

And from equation 1

$$\frac{(E_0)2}{(E_0)1} = \frac{(R_c)1}{(R_c)2} \text{ for } \begin{cases} |V| \le 1.5 \\ \dfrac{\omega_1 L_c}{R_c} \le 0.16 \\ \text{other parameters of equation 1 held constant} \end{cases} \quad (5)$$

Equations 4 and 5 may be used to establish the required change in cup re-

Table I

Item	Volts	Rpm	Present Aluminum Drag Cup*	Improved Aluminum Silicon-Bronze Drag Cup†
Relative air gap reluctance R_c			Relative air gap reluctance Rl	
Reactance / Resistance of drag cup $\left(\frac{\omega_1 L_c}{R_c}\right)$			0.14	0.04
Relative resistance of cup R_c			1	3.5
Cup temperature coefficient of resistance			+.24%/C (approximately)	+.06%/C (approximately)
Specific gravity of cup material			2.77	8.0 (approximately)
Output volts at 1,000 rpm (open circuit)			2.8	0.8
Relatively linearity error (ϵ_L)			−12	−1
$\pm\left\|\frac{\epsilon_L}{3}\right\|$ Maximum linearity error with speed compensation	0.3..2,400		±0.059%	±0.0049%
	0.4..3,200		±0.096%	±0.008%
	0.6..4,800		±0.22%	±0.018%
	0.8..6,400		±0.39%	±0.032%
	1.0..8,000		±0.61%	±0.051%

* Alcoa no. 24ST aluminum composition Al-93%, Cu-4.5%, Mg-1.5%, Mn-0.6%.

† Composition: Cu-85%, Al-10%, Mn-1% (representative only).

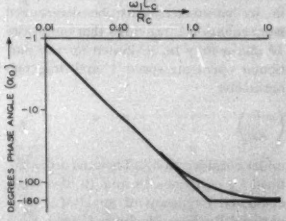


Fig. 8. Tachometer phase angle α_E as $V \to 0$ versus drag-cup parameter $\left(\frac{\omega_1 L_c}{R_c}\right)$

Note: Zero speed quadrature voltage resulting from iron losses has been neglected

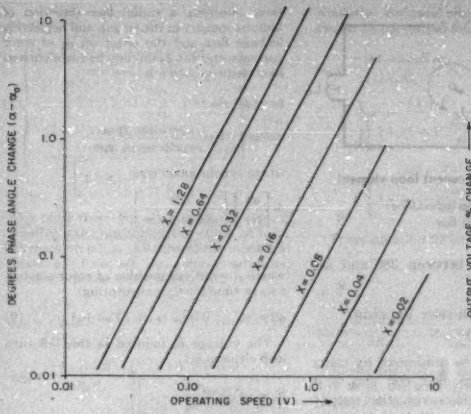


Fig. 9. Drag cup a-c tachometer degrees phase angle change $\alpha-\alpha_0$ versus operating speed V

$$X=\frac{\omega_1 L_c}{R_c}=\frac{\text{reactance}}{\text{resistance}}\ \text{of drag cup}$$

Note: Zero speed quadrature voltage has been neglected

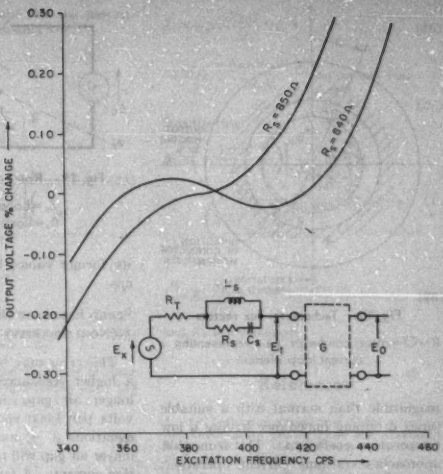


Fig. 10. Tachometer output voltage versus excitation frequency

$E_X = 230/-30°$ supply volts
$E_1 = 101$ volts excitation
$E_0 = 0.7/0°$ volts output at 1,000 rpm
G = motor tachometer with silicon-bronze drag cup
$R_T = 500\Omega$ = magnitude trim
R_s = frequency trim
$L_s = 0.500$ henry, $Q = 2.91$ @ 400 CY
$C_s = 0.15$ microfarad

sistance for the proper trade-off between linearity error ϵ_L and output volts E_0. Table I shows that if cup resistance R_c is increased by a factor of 3.5, then output volts per 1,000 rpm is decreased by this same factor, while linearity error is decreased by this factor squared.

The importance of using a cup material with a low temperature coefficient of resistance is shown in equation 5. Also, a low specific gravity material is desirable to keep the cup moment of inertia (integrator acceleration error) low.

The maximum linearity errors: $\epsilon_L/3$ from a median linear output line, realizable up to various speeds by the use of a proper scale factor (speed compensation), are shown in Table I for the two cups' materials tested. Fig. 8 shows output phase angle α_0 as operating speed V approaches zero versus cup parameter

$$\frac{\omega_1 L_c}{R_c}$$

This curve indicates the required change in phase angle of excitation voltage E_1 to keep the tachometer output phase angle constant as the cup parameter

$$\frac{\omega_1 L_c}{R_c}$$

is varied; see Fig. 3 and equation 1.

Fig. 9 shows the change in output phase angle $(\alpha-\alpha_0)$ versus operating speed V as a family of curves, each representing a particular value of cup parameter

$$\left(\frac{\omega_1 L_c}{R_c}\right)$$

Given the tachometer output volts/1,000 rpm, equation 1 and these curves, the change in quadrature voltage for a change in operating speed may be determined. Zero speed quadrature voltage resulting from iron losses has been neglected.

Conclusions

General

For this discussion refer to equations 19 and 1 and Fig. 3.

Output voltage E_0 is inversely proportional to the magnetic circuit reluctance R_g and to the drag cup resistance R_c. Maximum air gap length is limited by the maximum permissible heating and current in the excitation winding and the tolerable leakage flux.

A lower ratio $[\omega_1 L_c/R_c]$ from a higher cup resistance R_c and/or a longer air gap:

1. Reduces the time phase lag of output voltage α_0; see Fig. 5.
2. Reduces the time phase variation of output voltage $(\alpha-\alpha_0)$ with speed changes see Fig. 9.
3. Reduces the magnitude of output voltage E_0. See equations 1 and 5.
4. Reduces the nonlinearity error of output volts $(V-V_0)$ versus speed V. See Figs. 5 and 6 and equations 3 and 4.

The objects of this analysis were essentially to reduce by a specified amount: 1. the temperature coefficient of tachometer output voltage E_0, and 2. the tachometer speed linearity error of the output voltage E_0.

Since the component of drag cup magnetomotive force directly opposing the excitation winding magnetomotive force has been neglected, the phase shift data shown in Figs. 8 and 9 may be in considerable error. However, these curves should be useful to indicate trends. They are included for this reason.

Temperature Coefficient of Output Voltage

The change in value of stator flux ϕ_1 from changes in the excitation winding copper resistance may be minimized by using a voltage excitation source of higher

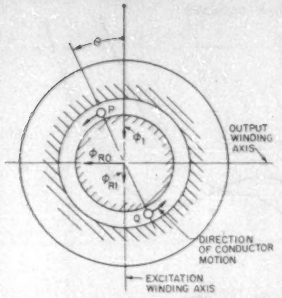


Fig. 11. Tachometer flux vectors

P−Q = rotor conductor pair representing a current loop element

magnitude than normal with a suitable series dropping impedance having a low temperature coefficient. This somewhat approaches constant current operation; see equation 1.

Larger wire may be used for the excitation winding to reduce its copper resistance and then add external resistance of low temperature coefficient to obtain the required phase angle of output voltage.

Drag-cup resistance R_c change may be minimized by using a material such as silicon bronze having a low temperature coefficient of resistance in place of the usual aluminum drag cup; see equation 5. By this means, the temperature coefficient of output voltage was reduced to −0.06% C, 1/4 of its former value.

The effect on output voltage E_0 from output winding copper resistance changes may be minimized by removing any output load trimming resistor and operating into a high load impedance, e.g., 1 megohm. Performance with the output winding loaded is not considered in this paper.

Frequency Coefficient of Output Voltage

With the compensation circuit shown in Fig. 10 (determined empirically) and for variations in excitation frequency ($f_1 \rightarrow f_2$), the corresponding variation of output voltage magnitude ΔE_0 was as follows:

ΔE_0	R_s	$(f_1 \rightarrow f_2)$
±0.05%	850	364–403
±0.03*	840	354–426
±0.20	800	410–430

* Interpolated curve.

Therefore, with the optimum circuit parameters, the variation of output voltage should be less than ±0.06% (1/10 of its former value) between 380 and 420 cps.

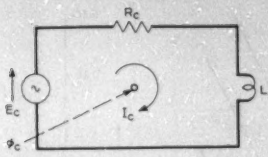


Fig. 12. Rotor current loop element

Z_c = Loop impedance
Φ_c = Loop flux

Speed Error of Output Voltage (Nonlinearity)

This error may be minimized by using a higher resistance drag cup and/or a longer air gap. However, the output volts per 1,000 rpm will be lower; see equations 3, 4, and 5. In addition, a longer air gap will result in higher excitation current. A family of linearity error curves may be plotted as shown in Fig. 6 using equations 19 and 1 with assumed values for

$$\frac{\omega_1 L_c}{R_c}$$

The curve matching the actual tachometer performance indicates the effective design parameter

$$\frac{\omega_1 L_c}{R_c}$$

By increasing the cup resistance R_c, by the substitution of a silicon bronze drag cup in place of the aluminum drag cup, the speed linearity error from a median linear output line was reduced to ±.05% up to synchronous speed (1/12 its former error) with an output of 0.8 volts per 1,000 rpm (1/3.5 its former value). Measured and theoretical data are in close agreement in this analysis where the tachometer output winding is not loaded.

Appendix

The basic elements of a two-pole tachometer are shown in Fig. 11 where:

ϕ_1 = excitation flux in maxwells
ϕ_{R0} = rotor flux along output winding axis
ϕ_{R1} = rotor flux along excitation winding axis
θ = radians position of rotor conductors representing a particular current loop element

let:

$$\phi_1 = |\phi_1|_{max} \cos(\omega_1 t) \tag{6}$$

then assuming a radial flux direction of uniform density in the air gap and neglecting leakage flux and the effect on ϕ_1 of rotor currents, the flux ϕ_2 linking the rotor current loop element shown is:

$$\phi_2 = (\phi_1) \sin(\theta) \tag{7}$$

define: $V = \dfrac{\text{operating rpm}}{\text{synchronous rpm}}$

where synchronous rpm

$$= \left[\frac{\omega_1}{2\pi}\right]\left[\frac{120}{\text{no. of poles}}\right]$$

Now: $\theta = (V\omega_1 t + A)$ (8)

where A = radians position of rotor conductors at time ($t = 0$), substituting:

$$\phi 2 = |\phi_1|_{max} \cos(\omega_1 t) \sin(V\omega_1 t + A) \tag{9}$$

The voltage E_c induced in this one-turn loop element is:

$$E_c = -(10^{-8})\frac{d\phi 2}{dt}$$

$$E_c = -|\phi_1|_{max}(\omega_1)\left(\frac{10^{-8}}{2}\right)\times \begin{Bmatrix}(1+V)\cos[\omega_1 t(1+V)+A]+\\ -(1-V)\cos[\omega_1 t(1-V)-A]\end{Bmatrix} \tag{10}$$

From Fig. 12 the current i_c in this one-turn element of the rotor is

$$i_c = \frac{E_c}{Z_c} \tag{11}$$

In the centimeter-gram-second system of units used here

$$F = \frac{4\pi}{10} Ni = \phi R_e \tag{12}$$

where:

F = mmf in gilberts
Ni = amperes × turns
$R_e = \Sigma S/\mu A$ = reluctance of air gap plus iron path
S = cm length of magnetic path
A = cm² cross section of magnetic path
μ = permeability of medium

now: $$\phi_c = \left(\frac{4\pi}{10R_e}\right) i_c \tag{13}$$

The component of elemental flux ϕ_{co} along the output winding axis is:

$$\phi_{co} = \phi_c \cos(\theta) = \left(\frac{4\pi}{10R_e}\right)\cos(V\omega_1 t + A)|i_c|$$

$$\phi_{co} = \left(\frac{-4\pi 10^{-8}\omega_1}{20R_e}\right)|\phi_1|\cos(V\omega_1 t + A)\times \begin{Bmatrix}\dfrac{(1+V)\cos[\omega_1 t(1+V)+A]}{R_e\left[1+j(1+V)\dfrac{\omega_1 L_c}{R_e}\right]}\\ \dfrac{-(1-V)\cos[\omega_1 t(1-V)-A]}{R_e\left[1+j(1-V)\dfrac{\omega_1 L_c}{R_e}\right]}\end{Bmatrix} \tag{14}$$

let:

$$K_1 = \frac{4\pi 10^{-8}}{20} \tag{15}$$

$$\phi_{co}=\frac{-|\phi 1|\max \omega_1 K_1}{2R_sR_c}\times$$

$$\left\{\begin{array}{l}\dfrac{(1+V)\cos[\omega_1 t]+(1+V)\times\cos[\omega_1 t(1+2V)+2A]}{\left[1+j(1+V)\dfrac{\omega_1 L_c}{R_c}\right]}+\\ \dfrac{-(1-V)\cos[\omega_1 t(1-2V)-2A]-(1-V)\cos[\omega_1 t]}{\left[1+j(1-V)\dfrac{\omega_1 L_c}{R_c}\right]}\end{array}\right\} \quad (16)$$

The total rotor flux ϕ_{RO} along the output winding axis, neglecting the effect of rotor currents on $\phi 1$, is the sum of the elemental fluxes ϕ_{co} from all the rotor conductors; i.e., from position ($0<A<\pi$).

$$\phi_{RO}=\int_{A=0}^{A=\pi}\phi_{co}dA \quad (17)$$

$$\phi_{RO}=\frac{-\theta_{1\max}\omega_1 K_1\pi}{2R_sR_c}\left\{\frac{(1+V\cos(\omega_1 t)}{\left[1+j(1+V)\dfrac{\omega_1 L_c}{R_c}\right]}+\frac{-(1-V)\cos(\omega_1 t)}{\left[1+j(1-V)\dfrac{\omega_1 L_c}{R_c}\right]}\right\} \quad (18)$$

From equation 18 let

$$\{\ldots\ldots\ldots\}=[(M)/+\underline{\alpha}]\cos(\omega_1 t) \quad (19)$$

then

$$\phi_{RO}=\frac{-\omega_1 K_1\pi(M)}{2R_sR_c}|\phi 1|\max_{\text{max}}\cos(\omega_1 t+\alpha) \quad (20)$$

Similarly, the total rotor flux ϕ_{R1} perpendicular to the output winding axis:

$$\phi_{R1}=\frac{-|\phi 1|_{\max}\omega_1 K_1\pi}{2R_sR_c}\times$$

$$\left\{\begin{array}{l}\dfrac{-(1+V)}{\left[1+j(1+V)\dfrac{\omega_1 L_c}{R_c}\right]}+\\ \dfrac{-(1-V)}{\left[1+j(1-V)\dfrac{\omega_1 L_c}{R_c}\right]}\end{array}\right\}\sin(\omega_1 t) \quad (21)$$

From equation 21 let

$$\{\ldots\ldots\ldots\}=(P)/\underline{+\beta} \quad (22)$$

then

$$\phi_{R1}=\frac{-\omega_1 K_1\pi(P)}{2R_sR_c}|\phi_1|\max_{\text{max}}\sin(\omega_1 t+\beta) \quad (23)$$

Neglecting leakage and the effect or rotor flux along the excitation winding axis, the air gap flux ϕ_1 in cgs units is

$$\phi_1\simeq\frac{4\pi N_1(i_1)_0}{10R_s} \quad (24)$$

where: $N_1(i_1)_0$ = excitation ampere turns with drag cup removed.
From Fig. 1

$$(i_1)_0=\frac{-jE_1}{\omega_1 L_1\left(1-j\dfrac{R_1}{\omega_1 L_1}\right)} \quad (25)$$

The output voltage is

$$E_0\simeq -N_2\frac{d(\phi_{RO})}{dt}\ 10^{-8} \quad (26)$$

$$E_0\simeq(j)\frac{-N_2\omega_1 K_1\pi(M)}{2R_sR_c}|\phi_1|\max_{\text{max}}\cos(\omega_1 t+\alpha) \quad (27)$$

Substituting

$$E_0\simeq\left(\frac{-4\pi^2 10^{-18}}{K_3}\right)\left[\frac{N_2\omega_1}{N_1R_sR_c}\right]\times\frac{M}{\left(1-j\dfrac{R_1}{\omega_1 L_1}\right)}|E_1|\max\cos(\omega_1 t+\alpha) \quad (1)$$

where:

$$L_1=K_3\frac{N_1^2}{R_s} \quad (28)$$

and K_3 is a constant.

References

1. An Induction Generator Used with the All-Electric Automatic Pilot for Airplanes, Herbert Vickers. *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 182–85.

2. Analysis of the Drag-Cup A-C Tachometer by Means of 2-Phase Symmetrical Components, R. H. Frazier. *Ibid.*, vol. 70, 1951, pp. 1894–1906.

3. Performance Characteristics of the Induction Generator Tachometer, Sidney Davis. *Product Engineering*, New York, N. Y., Jan. 1953, pp. 168–78.

4. Drag-Cup A-C Tachometer with Constant Current Excitation, R. H. Frazier. *AIEE Transactions*, vol. 72, pt. II, July 1953, pp. 150–52.

Discussion

Gerald Weiss (Microwave Research Institute, Brooklyn, N. Y.): The author has presented an interesting paper on a subject on which literature is not too abundent. He is probably correct in stating that the equations in the references he cites are too complex to yield useful results. However, I wish to point out that the author appears to have overlooked the most important reference in the field, a reference which is widely considered as standard for the a-c tachometer, the equivalent circuit and performance equations first published in 1950 by R. L. Cosgriff.[1] The identical equivalent circuit and performance equations were also derived in 1947 by Sidney Davis at the Arma Corporation. Davis' original reports were never released for publication, however he has presented some of his results in reference 3 of the paper. Both Davis and Cosgriff recognized that the a-c tachometer is merely a special case of an induction machine in general, and of a 2-phase induction motor in particular. They were then able to take their analysis and equivalent circuit directly from the prior art, in this case from S. S. L. Chang's paper on the capacitor motor.[2]

Nomenclature

A = tachometer gradient, defined in equation 33
a = secondary-to-primary turns ratio
B = nonlinearity constant, defined in equation 34
C = nonlinearity constant, defined in equation 35
D = nonlinearity constant, defined in equation 36
E_t = tachometer input voltage
E_0 = tachometer open-circuit output voltage
F = nonlinearity constant, defined in equation 37
f = frequency of E_t, cps
$j=\sqrt{-1}$
n = rotor speed, rpm
p = number of poles
R_e = added primary series resistance external to tachometer
R_m = core loss resistance, referred to primary, ohms
R_p = primary series resistance, ohms
R_r = rotor resistance, referred to primary, ohms
R_s = secondary series resistance, referred to secondary, ohms
v = rotor speed as a fraction of synchronous speed
X_m = magnetizing reactance, referred to primary, ohms
X_p = primary leakage reactance, referred to primary, ohms
X_r = rotor leakage reactance, referred to primary, ohms
X_s = secondary leakage reactance, referred to secondary, ohms
Z_t = input impedance of tachometer
$Z_m=R_m+jX_m$ = mutual impedance
Z_o = output impedance of tachometer, referred to secondary
$Z_p=R_p+jX_p$ = primary leakage impedance, referred to primary
$Z_r=R_r+jX_r$ = rotor impedance referred to primary
$Z_s=R_s+jX_s$ = secondary impedance, referred to secondary

Note: "Primary" refers to the phase to which the excitation is applied; "Secondary" refers to the phase from which the output voltage is taken. Strictly speaking, the rotor is the secondary.

The afore-mentioned equivalent circuit is reproduced in Fig. 13. This is the circuit for the main space wave of magnetomotive force; space harmonics can be taken into account by superposition, but the resulting complication is hardly worthwhile. From the circuit the significant unknowns, open-circuit voltage, input impedance, and output impedance are obtained by standard circuit analysis

$$\frac{E_0}{E_t}=\frac{Av}{1+Bv^3} \quad (29)$$

$$Z_o=Z_s+\frac{a^3\dfrac{Z_mZ_r}{Z_m+Z_r}+v^3C}{1+v^3B} \quad (30)$$

$$Z_t=Z_p+\frac{\dfrac{Z_mZ_r}{Z_m+Z_r}+v^3D}{1+v^3F} \quad (31)$$

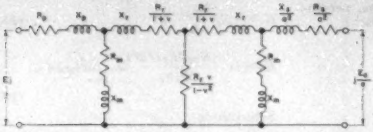


Fig. 13. Equivalent circuit of a-c tachometer

where the parameters v, A, B, C, D, and F are defined by

$$v=\frac{np}{120f} \tag{32}$$

$$A=\frac{-jaZ_m{}^2R_r}{(Z_m+Z_r)(Z_pZ_r+Z_pZ_m+Z_mZ_r)} \tag{33}$$

$$B=\frac{-(Z_m+Z_r-R_r)[(Z_pZ_r+Z_pZ_m+Z_mZ_r)-R_r(Z_p+Z_m)]}{(Z_m+Z_r)(Z_pZ_r+Z_pZ_m+Z_mZ_r)} \tag{34}$$

$$C=\frac{-a^2Z_m(Z_r-R_r)[(Z_pZ_r+Z_pZ_m+Z_mZ_r)-R_r(Z_p+Z_m)]}{(Z_m+Z_r)(Z_pZ_r+Z_pZ_m+Z_mZ_r)} \tag{35}$$

$$D=\frac{-Z_m(Z_r-R_r)\left[\frac{Z_s}{a^2}Z_r+\frac{Z_s}{a^2}Z_m+Z_mZ_r-R_r(Z_p+Z_m)\right]}{(Z_m+Z_r)\left(\frac{Z_s}{a^2}Z_r+\frac{Z_s}{a^2}Z_m+Z_mZ_r\right)} \tag{36}$$

$$F=\frac{-\left(\frac{Z_s}{a^2}Z_r+\frac{Z_s}{a^2}Z_m+Z_mZ_r\right)(Z_m+Z_r-R_r)+\left(\frac{Z_s}{a^2}+Z_m\right)[R_r(Z_r-R_r)+R_rZ_m]}{(Z_m+Z_r)\left(\frac{Z_s}{a^2}Z_r+\frac{Z_s}{a^2}Z_m+Z_mZ_s\right)} \tag{37}$$

The complicated equations 33–37 can be simplified by two assumptions valid for the drag-cup-type construction, namely

$$Z_m\approx jX_m \text{ negligible core loss} \tag{38}$$

$$Z_r\approx R_r \text{ negligible rotor leakage} \tag{39}$$

The simplified equations become

$$A=\frac{-ja}{\left(1+\frac{R_r}{jX_m}\right)\left(1+\frac{Z_p}{R_r}+\frac{Z_p}{jX_m}\right)} \tag{40}$$

$$B=\frac{-1}{\left(1+\frac{R_r}{jX_m}\right)\left(1+\frac{R_r}{jX_m}+\frac{R_r}{Z_p}\right)} \tag{41}$$

$$C=D=0 \tag{42}$$

$$F=\frac{-1}{\left(1+\frac{R_r}{jX_m}\right)\left(1+\frac{R_r}{jX_m}+\frac{a^2R_r}{Z_s}\right)} \tag{43}$$

The previous equations are capable of giving all the required information about gradient, phase shift, nonlinearity, temperature sensitivity, loading, and so on. This is now demonstrated by deriving some of the author's results.

For example, the condition of zero phase shift for the open circuit voltage can be obtained by setting the imaginary part of A equal to zero, obtaining

$$\frac{R_p}{X_m}=\frac{1+2\frac{X_p}{X_m}}{\frac{R_r}{X_m}-\frac{X_m}{R_r}} \tag{44}$$

Since $X_p<X_m<R_r$

$$R_p\approx\frac{X_m{}^2}{R_r} \tag{45}$$

If the actual primary resistance is smaller than the value obtained from equations 44 or 45, it is possible to add external series resistance R_g to obtain zero phase shift. As the author indicates in his conclusions, the added resistor can be fabricated of negative temperature coefficient material to counteract the positive temperature coefficient of the copper winding, and thereby obtain both temperature compensation and phase shift. (This design has actually been used as early as 1946.)

For this design the output gradient A becomes

$$A=\frac{a}{\frac{R_r+2(R_p+R_g)}{X_m}+\frac{X_p}{X_m}\left(\frac{R_r}{X_m}-\frac{X_m}{R_r}\right)} \tag{46}$$

and again for $X_p<X_m<R_r$

$$A\approx\frac{aX_m}{R_r+2(R_p+R_g)} \tag{47}$$

Similarly, equation 41 can be broken up into real and imaginary parts to obtain the direct and quadrature components of the linearity error. Comparison of equations 47 and 41 reveals that, to a first approximation, the output gradient varies inversely as R_r, while the lincarity error varies inversely as R_r squared. This is confirmed by the author's data.

It is this writer's opinion that the work of Cosgriff and Davis on the a-c tachometer is to be considered as definitive. It is based on the fact that this apparently mysterious device is nothing but a 2-phase machine used in a special way, namely with single-phase excitation and driven rotor. The various special features we encounter, such as drag-cup rotor, inside-out construction, and so on, were selected toward the end of obtaining special results; they do not alter the basic character of the device. It ought not to be necessary to introduce different methods of analysis, and different nomenclature, merely because an induction machine is used in a feedback control system rather than in a household fan.

It would be interesting to learn in what way the author's methods differ from those used before; whether they involve a set of approximations different than the conventional one of neglecting space harmonics and saturation; whether they lead to a different and possibly more useful equivalent circuit and performance equation; whether they lead to conclusions different from the ones obtained or obtainable from previous work.

References

1. Integral Controller for Use in Carrier-Type Servomechanisms, R. L. Cosgriff. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1379–83. (The analysis of the A-C tachometer is to be found in Appendix II of the paper.)

2. The Equivalent Circuit of the Capacitor Motor, Sheldon S. L. Chang. *Ibid.*, vol. 66, 1947, pp. 631–40.

L. A. Knox: Several remarks from the audience essentially involving compensating problems are discussed first.

Capacitive Coupling. A frequency sensitive residual output voltage will be present if there is appreciable capacitive coupling between the end turns of the excitation and output windings.

The tachometer output voltage versus excitation frequency for one direction of rotation is the sum of this frequency sensitive residual voltage and the speed generated voltage shown in Fig. 4. For the reverse direction of rotation, the tachometer output is the difference of these two voltages.

In order to allow compensation of the tachometer output voltage frequency coefficient for both directions of rotation, this capacitance effect should be minimized (by an electrostatic shield or equivalent).

Frequency Compensating Network. The author is indebted to Mr. H. Sabath of Servomechanisms, Inc. for the following comments (summarized by the author).

The R-L-C network of Fig. 10 represents an additional series impedance in the tachometer excitation circuit which causes poorer regulation of the tachometer generator. In this case, the load is represented by drag cup currents that increase with operating speed; see equations 18 and 21.

Therefore, output voltage speed linearity error is adversely affected by additional series impedance in the excitation circuit. These effects, including leakage reactance, neglected to simplify this approximate analysis, may be partially simulated in Fig. 3 by replacing $(R_1)_F$ with its parallel equivalent across L_1, adding a load resistor across L_1 to represent drag cup rotational currents, and adding leakage reactance in series with $(R_1)_e$.

The recommended method for frequency compensation involves a standard network in the velocity servo-amplifier input. Such a network is shown in Fig. 14 and may be

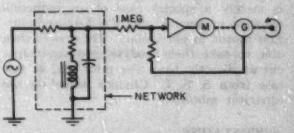


Fig. 14. Frequency compensation network

synthesized to match the tachometer output voltage versus excitation frequency characteristic; see Fig. 4.

I appreciate the time Mr. Weiss has taken to prepare his comprehensive discussion. Of special value are his equations 41 and 47 which on further simplification yield the results that, for a precision integrating tachometer, the output gradient varies inversely as cup resistance and linearity error varies inversely as cup resistance squared. This trade-off indicates that, in general, output gradient must be sacrificed to obtain a small linearity error.

My paper is a simplified analysis of the tachometer, written particularly for the user.

Only characteristics of primary interest to him are presented, and these are expressed as parameters that he can measure or readily interpret.

Practical Air Filtration for Diesel Locomotives

M. B. ADAMS
NONMEMBER AIEE

VARIOUS forms of dirt and deposits in the engine and electric equipment of a diesel-electric locomotive are generally considered harmful. The purpose of filtration equipment is to remove as much of the dirt as possible, within economic limits. There is, however, a certain level of cleanliness beyond which it is not feasible to operate.

Dirt and Its Effects

Electric rotating equipment on diesel-electric locomotives usually consists of d-c machines. The principal units, and those requiring the most maintenance, are the main generator, traction motors, and the various auxiliary generators and motors. Most of this equipment was initially placed in service on locomotives not equipped with any type of air filter. In spite of the fact that some of these units, principally switch-type diesels, operate in very dirty air conditions on all railroads, we have had relatively few difficulties due to the abrasive effect of dirt. Inspection of the commutators of motors and generators on these locomotives will rarely, if ever, reveal any indication that these parts are actually worn because of dirt. This is, no doubt, due largely to the centrifuge effect of the rotating armature which, in part, is self-cleaning.

In this type of equipment there is another form of dirt which is more harmful and that is the self-generated variety. The dirt, or carbon dust, which forms from the normal wear of commutator brushes can cause considerable trouble, especially when a binding agent is present, such as oil and/or water vapor. In such cases, the dust will adhere to the damp surface, and carbon paths can be and are formed across insulated areas, resulting in short circuits and grounded circuits. These conditions are often serious and result in failure of the equipment in service, which is costly.

The more ordinary type of dirt causes more trouble in the control circuits and power switchgear than any other place in the electric equipment. Small particles of dust can cause a maximum of trouble completely out of proportion to the size of the particle. The electric control circuits can be quite comprehensive in nature in order to provide the proper degree of safety in locomotive operation. A small particle of dirt in any one of the interlocks in the system can cause a malfunction of the locomotive. Because of this, locomotives were designed with control circuit interlocks having adequate "wipe" to brush away any such particles of dirt and complete the necessary circuit.

In spite of this design, we were plagued with interruptions in service, and early failure of the control and power circuit interlocks and contactors because of this wipe. The dirt particles in unfiltered air acted as an abrasive and the wiping action provided the necessary mechanical force to create wearing action.

Air Filtration

This situation demanded the use of carbody air filtration in diesel locomotives and in 1941 it was decided to equip certain locomotives with carbody filters to filter all the air entering the carbody. These applications were first made on freight locomotives, reducing the engine room dirt conditions. However, the panel filter itself, as well as the design of the application, was subsequently improved. The original application had a slot below the filter panels to allow free flow of air into the carbody in the event filter panels plugged with dirt. New panels of a nonplugging design were developed and the slots were closed off to provide complete air filtration for all engine room air. Today applications of this type are standard equipment on all road and road switcher-type locomotives.

The improvement in the quality of air for the locomotive has also brought about newer designs in relays, interlocks, and contactors. Some of these new designs are good in that they are smaller, less expensive, and more trouble-free. While the dust remaining in the filtered air in the locomotive is now of a small particle size, it can still cause trouble. We therefore still need a good wiping action to assure good electric contact to complete the various circuits. We must also have surfaces that are not prone to cinder formation, such as is found on copper contact bars and fingers. The newer interlocks and some heavy duty contacts are silver plated or have a silver inlay in the mating portion of the contact tips to provide a surface that is not as susceptible to dirt and the formation of cinders.

In addition to carbody filters, there are some other designs of air filters that can be applied to main generators and traction motors. As has been previously mentioned, the presence of a binder causing dust or dirt particles to cling to insulated parts and creepage surfaces is a problem in these machines. Some attempts have been made to use the conventional 2-inch impingement filter as a dry element to stop oil and water vapors and dirt from entering the machine. Depending on the severity of the problem, this can be a productive application. However, it might be more practical to reduce or eliminate the vapor from the air.

The principal sources of moisture-carrying vapors in the engine room are the engine exhaust system, the steam generator, and the air compressor. By maintaining these items properly, most of the

Paper 58-500, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-ASME Railroad Conference, Cleveland, Ohio, April 9-10, 1958. Manuscript submitted December 26, 1957; made available for printing January 28, 1958.

M. B. Adams is with the Atchison, Topeka and Santa Fe Railway, Chicago, Ill.

moisture can be kept out of the electric equipment, and what dust does get into the generators and motors will either be blown out or will be relatively harmless.

Economic Considerations

Main Generator

One particular problem with main generators on the Santa Fe Railway has been evaluated through test applications of impingement-type filters. We have found minor differences in the dirt deposits between those generators equipped with filters and those not so equipped. However, we have never been able to detect any difference in the over-all operation. The difficulties experienced were unrelated to the use of filters. The filters were used both dry, and coated with conventional plasticized filter saturant. The use of filters brought out two unexpected problems; small pieces of the filter would break off and mark the main generator commutator, and pieces of dirt would drop off the filter into the generator and damage the brushes and commutator. It was also found that excessive oiling of the filter could result in the saturant being pulled into the main generator to collect dirt there, especially at higher ambient engine-room temperatures such as are experienced in our desert operation.

Because of these things, the test was terminated with the conclusion that the cost of the application could not be offset by measurable savings. It should be said that other railroads with problems of a different or more severe nature have found such filtration desirable; however, if at all possible, the more basic approach of elimination of the causes is preferable.

Traction Motor

Diesel units equipped with steam generators are prone to experience more electrical difficulties arising from moisture grounds. Statistics show that 45.8% of the traction motor grounds or short circuits, on a unit equipped with a steam generator, occur in the no. 4 traction motor, which is directly below the steam generator. The incipient steam and water leaks, occurring in such equipment, not only result in moisture draining onto the traction motor; but in some cases also, steam leaks, adjacent to the traction-motor blowers, are found in the air stream which flows into the blower. The steam is drawn into the motor and the moisture condenses on the brush box insulating studs. This can and does result in grounded traction motors, ground relay action, and operational failures.

The dirt conditions in traction motors are, of course, much worse than in generators because of their location; hence the presence of moisture aggravates the possibility of grounds. To prevent the entrance of moisture into the traction motors, designs are now available to extend the point of air intake to a higher location in the carbody. This will prevent steam from being drawn into the motors. However, other steps must be taken to seal the traction-motor air duct opening properly against the entrance of water, etc., that may be found on the engine room floor.

Filtration Methods

A representative diesel engine will require approximately 15,000 cubic feet per minute of air during full-load operation. The use of carbody filters has improved locomotive performance to a great extent but not without disadvantages. The principal difficulty is that a negative pressure is produced within the carbody. This means that any small hole, even drain holes, will actually become a point of admission of unfiltered air. During snow and rain storms moisture is drawn into the carbody; and in dry climates, dust can also enter. During winter months it is sometimes necessary to block certain filters to prevent entrance of blowing snow. This increases the pressure differential and can cause more snow to be drawn into the carbody. As a result, we must make some concessions to actual conditions and find it necessary to apply special traction motor air ducts and special insulating materials.

The real solution could be the complete pressurization of the carbody by the use of a skimmer-type blower, or by some other method. The Fairbanks-Morse Company has built three locomotives equipped with this type of equipment for one of the Eastern railways. One of the units, as previously mentioned, used the skimmer-type blower. The other two units use large banks of Rotonamic filters which provide clean air for engine consumption in sufficient quantities so that a slight pressurization of the carbody is achieved. To date, it is not possible to quote any actual operating statistics that might indicate the efficacy of this application.

These applications will, in time, provide many of the answers now being sought regarding the benefits of filtered air to electric equipment. The basic idea is good, and could be justified, not only by reduced maintenance costs, but also by better performance. The skimmer blower is a roof-mounted blower drawing in outside air through a roof opening. It is a squirrel-cage-type blower that by-passes part of the air it draws into the carbody. The dust particles, being heavier, are spun to the outside of the discharge opening of the blower, where they are skimmed off and discharged outside the cab. The clean air is admitted directly to the engine room. The capacity of the blower being greater than the air requirements of the locomotive enables it slightly to pressurize the carbody. This prevents dirt and moisture from being drawn into the engine room and electric equipment. There are no carbody filters or other openings by which air can enter the engine room.

A similar idea is used in the winterization hatch, which is furnished by the Electro-Motive Division (EMD) of General Motors. However, this hatch utilizes existing equipment on the locomotive and can allow unfiltered air to enter the engine room. EMD has also used a "snorkle" design for locomotives built primarily for export to the Middle East. This particular application provides for full filtration of engine-room air and exclusion of sand, dust, and other air-borne contaminants.

Because of the design problems and cost of application to existing equipment, systems such as described are best applied to new locomotives while in production and it is more practical for the locomotive manufacturers to take the lead in this type of locomotive improvement.

Control Cabinet

In 1953 a "sealed cabinet" project was initiated on the Santa Fe Railway with one of our, then, new *F-7*-type locomotives. The electric cabinets were to be operated from monthly inspection to monthly inspection without the need for being opened and inspected. The equipment included new cam switch contacts; new enclosed motor shunting contactors; new parallel relay, battery field, and shunt field relays; and a new battery charging contactor. These relays and contactors, together with the power contactors, were equipped with new, enclosed, interlock assemblies. This project ran for about 3 years, during which time several difficulties were encountered and resultant changes made. The initial difficulty was the matter of education of maintenance and operating people to respect the seals on the electric cabinet doors. Eventually, our personnel did respect the seals and quite good performance resulted. Much of the equipment to come out of this test has since become standard. Even so, we were never able to operate for extended periods of time without "blowing" the cabinets. It was

always necessary to blow the dust out of the cabinets, clean and lubricate the reverser, and make the routine inspection for visible defects. In some cases it was necessary to dismantle and clean the sealed interlock assemblies, as we found them to suffer from dirt particles which found their way into the assembly. In spite of the improvements to come out of the test, we have not been able to dispense with basic electrical maintenance and periodic cleaning.

It is still necessary to blow electric rotating equipment, wipe insulators, and blow dirt out of the electric cabinets. Of course, the work is easier; but it still must be done to provide the necessary insurance against failures. For this reason, it is difficult to equate improved filtration to a reduction in maintenance cost of electric equipment.

Diesel Engine

The evaluation of the cost of dirt elimination in diesel engines can be more definitely explained. Wear of the internal parts in a diesel engine is a direct result of dirt getting into the engine. Of course, there are some variables, such as the condition in which the lubricating oil is maintained and average engine temperatures. Low engine-cooling water temperature can result in high wear rates. However, these undesirables are the exception rather than the rule. Before the importance of air filtration was realized, diesel switch locomotives were operating one year between cylinder assembly change-outs. Today these same engines operate a minimum of 2 years, and in some cases longer, between assembly renewals. The principal difference is that of improved engine air filtration. The old Burgress snubber and air filter has been replaced by the panel-type impinement filter, which is more efficient and easier to handle.

The conventional 2-inch impingement-type panel has been replaced with a 4-inch panel-type element in an effort to provide better filtration and extend the necessary change-out periods. While this filter does do a better job, it also emphasizes the necessity for a filter that will provide a uniformly excellent level of filtration. We have found such performance from the Rotonamic-type filter and are using it on nearly all applications of new filters. The 4-inch filters that are replaced by the Rotonamic are being applied to yard switch engines to give them better protection, while the 2-inch filters replaced in this program are being used in carbody filter applications.

In much of our territory on the Santa Fe Railway, from central Kansas, south through Oklahoma to Texas, and west through Colorado and New Mexico, very severe dust conditions exist. We find the impingement filter overloaded after only a few hours of operation and daily changes of carbody and engine air filters are required. Such frequent filter changes, of course, are impossible, both from the standpoint of maintaining clean filters ready for application and of having the locomotive available for filter changes. This results in severe dirt conditions in the engine. During severe dust storms, the impingement-type filters usually become dry and allow dirt to enter the engine, causing excessive ring and linear wear. Such a condition can arise even though the correct filter saturant is being used in the proper quantity, because the oil cannot wet through the collecting dust as fast as additional dust is presented with the intake air.

In the areas where these conditions exist, we have found it both beneficial and expedient to use the Rotonamic filter, which is a self-cleaning element using the "centrifuge" principle, to clean the air. It is quite efficient when handling dirt sizes above the 5-micron range. Here again our tests indicate this to be a critical size. Particles below this size will not contribute appreciably to engine wear. However, particle size greater than 5 microns definitely does increase engine wear.

Tests have been conducted to compare engine linear wear in engines using the Rotonamic filter with those using the 4-inch impingement filter. These tests indicate that the 4-inch filter will do a good job as long as it is well maintained and not allowed to become overloaded or dry. Cylinder liner wear will be approximately 0.011 inch at the point of maximum ring travel. This wear occurs in a 2-year period of heavy freight service during which time the unit travels approximately 224,450 miles. In this period the locomotive unit has had about 40 filter changes of 16 filters each. It has been determined that it costs about 37 cents per filter to remove, clean, re-oil, and reapply. This cost, equated to the annual mileage of the unit, gives a filter maintenance cost of 0.105 cent per mile.

Cost Comparisons

Diesel Engine

When establishing a basis for calculation of the cost of engine maintenance as compared to the cost of air filtration, it must be remembered that there are many variables combining to produce engine wear. Not the least of these are weather, temperatures, geographic conditions, etc. In a damp tropical climate very little airborne dirt would be presented to the engine or the filters. In areas such as we have in the "dust bowl" great quantities of abrasives must be filtered out of engine air. From experience with various methods of air filtration and tests of diesel engine wear patterns, it seems acceptable to estimate a 50% reduction in engine wear because of good air filtration. Conversely, a total absence of engine air filtration would require assembly renewal in EMD engines yearly instead of every 2 years as is done at present. During this time half of the liner life would be lost, piston rings would be worn out and, in all probability, exhaust valves would require replacement.

Considering the 16-cylinder, *567-B* EMD engine, our cost of operation would be as follows:

Cost of parts	
16×1/2 (cost of liner) =	
16×1/2×$95.00 =	$ 760.00
16×(cost of ring set) =	
16×$5.16 =	82.56
16×4 (cost of exhaust valve) =	
16×4×$4.50 =	288.00
16×(cost of gasket set) =	
16×$3.58 =	57.28
Total =	$1,187.84
Cost of application of parts =	230.21
Additional cost of parts and labor over 2-year period =	$1,418.05
Additional annual cost = 1,418.05 =	$ 709.02
Less annual cost of filter maintenance =	118.40
Annual cost increase due to elimination of filters =	$ 590.62
Increase cost per unit mile $ 590.62 = 0.53 cent per mile 112,225	

This assumed cost is not the most satisfactory method of determining the savings from air filtration. However, the only way to improve upon it is actually to test the engine on a diet of unfiltered air. This is not likely to happen, as no one today could afford to take the risk or pay the cost. Actually, additional costs should be considered, such as availability, interruption in service, increased handling charges, etc. These are really not necessary for this purpose, for we are interested in tangibles only. The charges mentioned would only increase the already prohibitive cost of unfiltered air, for engine air filtration has a direct bearing not only on liner, valve, and ring wear, but also on lubricating oil condition. Dirt entering the engine because of poor engine air filtration finds its way into the lubricating oil, causes wear of other

engine parts, and throws an increased load on the lubricating oil filters. This points up the fact that the filtration of air for combustion, and the filtration of lubricating oil and fuel oil, are the most important items of routine maintenance in the diesel locomotive.

Electric Equipment

In considering the cost of air filtration for the electric equipment, the total yearly cost of filter maintenance of $118.40 is reduced by that portion applying to engine air filters. This results in a cost of carbody air filtration of $76.96 per year for protection of electric switchgear and the various control relays. The corresponding cost per mile for this protection is 0.0685 cent.

The direct manner of evaluation used in the example of diesel engine wear versus filter cost relationship is not readily applicable to the electric system. Electrical failures due to dirt in contacts or interlocks can be quite costly, as it might often be necessary to send an electrician or some other properly advised craftsman to make corrections at the point of failure. Obviously cleaner air is sure to result in fewer road delays and less worn or burned parts. Even so, it is a real problem to make a direct statement as to the number of road failures that would be eliminated or the extension of life to the various parts. The current prices of contact tips and interlock fingers for various parts is shown in the following table; these are for present EMD items and include the approximate cost of application:

Cam switch	$217.50
Reverser	360.00
Power contactors	114.50
Auxiliary relays	47.00
Total	$739.00

Since it is admittedly difficult to establish a normal life expectancy for these parts, we can only estimate what our savings may be. A minor saving annually of only 1/10th of the replacement cost would be an attractive savings.

It is regrettable that a more definite method of fixing costs of dirt in electric equipment is not available. In the absence of such information, it remains for more definite proof to be established.

An Evaluation of the Effect of Dirt on Diesel-Electric Locomotives

J. W. HORINE
MEMBER AIEE

THE EVALUATION of dirt costs on diesel-electric locomotives made in this paper is admittedly extremely approximate in nature because of the scarcity of factual data and the many assumptions used in lieu thereof. It divides the locomotive into three main categories: the diesel engine (with cost figures of fair accuracy), and electric apparatus and other apparatus with dirt costs of "order-of-magnitude" approximations only. Dirt costs are estimated on a total national basis, divided between cleaning and repair costs, with further subdivision as avoidable or nonavoidable.

Physical Characteristics

While dirt on railroad equipment has always been a problem, its deleterious effects on diesel-electric locomotives are particularly serious, and the two operations of cleaning and repairs are interdependent. For example, failure over a period of time to correct an oil leak can materially increase cleaning cost in order to comply with ICC (Interstate Commerce Commission) regulations; on the other hand, failure to keep engine rooms clean will result in excessive dirt in engine air with corresponding increased wear rates on expensive engine components.

The generic term "dirt," in so far as a diesel-electric locomotive is concerned, includes three main categories: 1. sand and roadside dirt, abrasive in nature, 2. brake shoe dust, which is not only abrasive but also partially conductive and magnetic, and 3. trash, such as cigarette package and lunch papers, leaves, and oily and dirty rags.

Location and Effect of Dirt. The locations where this dirt is found, and its ill effects, can be summarized as in Table I.

Economic Aspects

Cost of Removal. One of the major dollar losses resulting from dirt is the cost of its removal. A survey at a major diesel shop, equipped with a power-driven trackside washer, developed a schedule of costs per unit as shown in Table II.

When these figures in Table II are applied on a national basis, the annual cleaning bill to the railroads of the United States is $39,717,211; the details are shown in the Appendix.

Diesel Engine. Another major cost of dirt is that resulting from its entry into the diesel engine itself. This cost is susceptible to reasonable evaluation. A filter efficiency test, using radio-activated rings to measure wear rate, developed the filter performance figures shown in Table III.

From Table III, we find that the wear rate of the piston rings on engines equipped with 2-inch standard oil-coated impingement-type filter is 1.412/0.402 or 3.51 times as great as with a type *A* or *B* filter.

Further, over a period of 3 years and 10 months, service tests were made on a group of units with type *A* and *B* engine air filters versus comparison units having 2-inch standard oil-coated impingement filters. The comparative material costs

Table I. Location and Effect of Dirt

	Location	Effect
A.	Surface Dirt	
	Exterior of car body	Poor appearance and finish damage
	Trucks and running gear	Fire hazard and wear
	Top of fuel oil tanks	Fire hazard
	Windows and windshields	Crew annoyance and safety hazards
	Interior walls, floors, and ceilings	Fire and slipping hazard
B.	Internal Dirt	
	Traction motors*	Flashovers, shorts, and grounds
	Main generators*	Flashovers, shorts, and grounds
	Auxiliary rotating electric apparatus*	Flashovers, shorts, and grounds
	Control cabinets	Malfunctioning of controls and fire damage
	Diesel engines†	Excessive wear rates

* Also generate own dirt through silica and carbonaceous scourings from brushes.
† Also generate own dirt through blow-by and oxidized oil.

Paper 58-504, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-ASME Railroad Conference, Cleveland, Ohio, April 9-10, 1958. Manuscript submitted December 13, 1957; made available for printing February 13, 1958.

J. W. Horine is with The Pennsylvania Railroad Company, Philadelphia, Pa.

Table II. Cleaning Costs

Light (interim) cleaning:	
Labor	$ 3.64
Materials	1.84
Total	$ 5.48
Heavy (monthly) cleaning:	
Labor	$81.91
Materials	3.15
Total	$85.06

Table III. Filter Performance

Filter Type	Milligrams of Iron per Gram of Oil-Free Dust to Engine
A and B	0.402
C	0.029
D	0.956
2-inch impingement-type, lube oil coating	1.412

Table IV. Material Costs of Engine Components

	Cylinder Liners	Pistons	Connecting Rod Bearings	Main Bearings	Total
A and B filter units	$ 957.00	$ 470.80	$240.70	$600.60	$2,269.10
2-inch filter units	2,784.00	1,412.40	716.30	673.20	5,585.90
Ratio, 2 inches to A and B	2.91	3.00	2.98	1.12	2.46

Table V. Comparison of Material Costs

	Measure of				
Filter	Ring Wear	Liner Wear	Piston Wear	Connecting Rod Bearing Wear	Main Bearing Wear	Over-all Wear
2-inch oil coated	$ 1.412	$2,784.0	$1,412.4	$716.3	$673.2	$5,585.9
A or B	0.402	957.0	470.8	240.7	600.6	2,269.1
Per cent reduction	71.5	65.7	66.7	66.4	10.8	59.4

of the components charged out during this time, average per unit, are given in Table IV.

The two tests of Tables III and IV are summarized in Table V. From this table we find that the reduction per unit in material cost only is:

$$\frac{\$5,585.90 - \$2,269.10}{\text{3 years, 10 months}} = \$865.25 \text{ per unit per year}$$

However, the savings from reduction of dirt in the diesel engine do not stop here. The period of time, or mileage interval, at which diesel-electric locomotives must be back-shopped for heavy or mileage overhauls is largely determined by the condition of the diesel engine, and any lengthening of this period produces substantial economies. This results not only from the greater service life between overhauls of the diesel engine itself, but also from the greater service life obtained from other components which must receive overhaul at the same shopping in order to avoid repeated shoppings. That is, a portion of the other components' potential service lives, available but not used on account of overhaul at the shorter period, is used at the long overhaul period.

A comparatively recent study, based on the then existing shopping periods and the reduced wear rates tabulated in Table V, indicated that shopping periods could be extended 6 months by use of *A* or *B* engine air filters with an average annual saving per unit of $313.17.

When these two factors are applied on a national basis, the annual avoidable cost of dirt in the diesel engine is estimated at $37,346,369. Again, details are in the Appendix.

Electric Apparatus. Unfortunately, the dirt damage to electric apparatus cannot be evaluated or even estimated with reasonable accuracy because of the multiplicity of causes of electrical failures, the lack of accurate analyses of these failures, and the paucity of detail cost data on repairs to electric apparatus. Further, the variable and unholy alliance of dirt and moisture, rather than the presence of either one alone, is undoubtedly a major factor.

As illustrative of the complexity of this subject, Table VI lists the more common electrical faults on diesel-electric locomotives and their possible causes. It will be appreciated that railroads cannot economically justify the manpower for the detailed analyses and cost accounting

Table VI. Diesel-Electric Locomotives

Common Electrical Faults and Their Causes

Fault/Cause	Dirt and/or Moisture	Excessive Gear Wear	Broken Brushes	Improper Grade of Brush	Improper Workmanship	Age or Wear	Improper Lubrication	Bearing Failure	Poor Design	Overloading Unit	Slipping Wheels or Overspeed	Brake Shoe Sparks	Loose Pinion	Malfunctioning of Controls
Armature winding shorted or grounded	X	X			X	X				X				X
Armature winding to be replaced	X	X			X	X		X		X	X		X	X
Armature bands damaged	X				X	X		X			X		X	
Commutator flashed	X	X	X	X	X				X	X	X		X	X
Commutator grooved		X	X	X	X	X		X	X	X				
Commutator loose					X	X		X	X		X		X	
Commutator to be renewed		X		X	X	X		X	X	X	X		X	X
Broken brushes	X	X		X				X			X		X	X
Brushholders electrically damaged	X	X	X	X	X			X			X		X	X
Brushholders mechanically damaged					X			X						
Field coils shorted	X	X			X	X		X		X				
Field coils grounded	X	X			X	X		X	X	X				
Field coils roasted					X	X			X	X				X
Field coils loose on pole		X			X			X	X					
Armature bearing damaged	X	X			X	X	X				X			
Loose pinion		X			X	X	X				X			X
Motor support excessive play		X			X			X						
Burnt wires and cables	X				X	X			X	X		X		X
Control and power grounds	X				X	X			X	X				
Malfunctioning of controls	X				X	X	X		X					
Control and power cabinet fires	X				X				X	X		X		X

that would be required to accurately evaluate each failure.

Undoubtedly, some of this dirt damage is avoidable. With present-day design, some additional filtration of main generator and traction motor cooling air is possible, and dust-proofing of control and power cabinets could be improved. Future units could have sealed and pressurized engine rooms with all incoming air filtered through type *A* and *B* filters, further reducing dirt content of air resulting in reduction in dirt damage.

For reasons previously stated no reasonably accurate estimate of savings to be expected from improved dirt elimination is possible, but an "order-of-magnitude" approximation indicates an annual national cost of $23,564,938. Again, details are given in the Appendix.

Additional Losses. There are many other dollar losses resulting from dirt which, while individually small, are very appreciable in the aggregate. Among these are heat transfer losses in radiators and heat exchangers, damage to the exterior finish on cab bodies, fire damage, and wear between moving surfaces such as pedestal jaws and boxes, brake and spring rigging pins and bushings, V-belt, chain-belt and geared drives, relay and contactor pins, etc.

Any evaluation of the cost of these losses is manifestly impracticable, nor are corrective measures readily apparent. This dirt loss must therefore be classed as unavoidable, and only "order-of-magnitude" approximation can be made. This approximation on a national basis is $38,104,867, as shown in the Appendix.

Summary

The presence of dirt on and in diesel-electric locomotives therefore results in out-of-pocket dollar expenditures not only for its removal (the cost of which is chargeable to Enginehouse Expenses, ICC Accounts Nos. 388 or 400, a Transportation Charge), but also the increased repair costs (chargeable to Locomotive Repairs, ICC Account No. 311, a Maintenance of Equipment charge) resulting from its presence. These expenditures are detailed in Tables VII and VIII.

In brief, it is estimated that dirt is presently responsible for approximately 25 per cent of the total cost of maintaining and preparing diesel-electric locomotives for service; and that slightly less than half of this present loss is avoidable. Whether or not the cost of avoiding this portion is warranted by the savings obtainable would require detailed study of each individual case.

Table VII. Estimates of National Dirt Costs on Diesel-Electric Locomotives

Item	Dirt Loss Unavoidable	Avoidable	Total
Cleaning	$39,717,211		$ 39,717,211
Diesel engines		$37,346,369	37,346,369
Electric apparatus		23,564,938	23,564,938
Other apparatus	38,104,867		38,104,867
Total	$77,822,078	$60,911,307	$138,733,385

Table VIII. Per Cent of National Total Costs

ICC Account No.	Name	National Total	Dirt Damage Estimate	Per Cent
388 and 400	Enginehouse expenses	$136,361,652	$ 39,717,211	29.1
311	Locomotive repairs	435,484,197	99,016,174	22.7
		$571,845,849	$138,733,385	24.3

Appendix. Data for Year 1956

Data:

Cost of interim cleaning of road units*	$ 5.48
Cost of monthly cleaning of road units*	85.06
Increased annual material cost of part changeouts of engine with 2-inch standard oil coated impingement air filters as compared to engines with type *A* or *B* filters*	865.25
Annual savings on units with type *A* or *B* filters as compared to units with 2-inch standard oil-coated impingement filters through extended overhaul period*	313.17

Units in service, United States and Canada:[1,2]

	No.	Per Cent
Over 1,200 hp (horsepower)	19,668	66.4
1,200 hp and under	9,951	33.6
Total	29,619	100.0

Units in service in United States:[3]	
At beginning of year	24,756
At close of year (average: 25,419)	26,081
Enginehouse Expenses—Yard[4]	$ 30,844,543
Enginehouse Expenses—Train	105,517,109
Total	$136,361,652
Cost of Locomotive Repairs—Yard[5]	$ 70,493,874
Cost of Locomotive Repairs—Other	364,990,323
Total	$435,484,197

Assumptions:

1. Division of units between "over 1,200 hp" and "1,200 hp and under" in the United States and Canada combined.

2. Units over 1,200 hp are road-type power and units 1,200 hp and under are yard switcher type.

3. Monthly cleaning of yard switcher locomotives costs only two-thirds that for road locomotives.

4. Road units receive interim cleaning every other day. Yard switcher units do not receive any interim cleaning chargeable to Enginehouse Expenses.

5. Engine air filters in service in 1956 were as follows:

Type	Per Cent	Milligrams of Iron per Gram of Dust*
A or B	10	0.402
C	25	0.629
D	25	0.956
2-inch standard oil coated impingement	40	1.412

* From test.

6. Dirt damage in the engine is directly proportional to the test milligrams or iron per gram of dust as developed by tests of four types of filters.

7. Total cost of repairs to diesel-electric locomotives are 60% labor and 40% material.

8. Total cost of repairs to diesel-electric locomotives are divided approximately as follows:

Diesel engine and appurtenances	35%
Electric apparatus	30%
All other	35%

9. Due to multiplicity of causes of electric faults, dirt damage on electric apparatus is of only half as much importance as in the diesel engine.

10. Elimination of dirt damage to other than the diesel engine would not extend the period between overhauls of the unit.

11. Dirt damage on those portions of the unit other than the diesel engine and electric apparatus is probably responsible for approximately a quarter of the cost of repairs to these components.

* From test results available to and previous studies made by the author.

Estimated Cleaning Costs:

Road units in service = 25,419×66.4% = 16,878
Yard units in service = 25,419×33.6% = 8,541
Total units in service = 25,419

Annual Cost of Cleaning:

Road units = ($5.48×15+$85.06)×16,878×12 = $33,876,171
Yard units = ($85.06×0.67)×8,541×12 = 5,841,040
Total annual cleaning cost = $39,717,211

This item is chargeable to Enginehouse Expenses, totalling $136,361,652, and is 29.1% thereof.

Estimated Diesel Engine Damage:

Avoidable part changeouts costs per unit: $\left($865.25+$865.25×$\frac{60}{40}$$\right)$ = $2,163.13
From extended service life between overhauls = 313.17
Total per unit, *A* or *B* filter versus 2-inch filter = $2,476.30

With allowance for estimated type of filter, national avoidable cost is:

$\frac{25,419×0.40×\$2,476.30×1.412-0.402}{1.412-0.402}$ = $25,178,028

$\frac{25,419×0.25×\$2,476.30×0.956-0.402}{1.412-0.402}$ = 8,631,576

$\frac{25,419×0.25×\$2,476.30×0.629-0.402}{1.412-0.402}$ = 3,536,765

Total probable annual avoidable damage = $37,346,369

Dirt Damage on Electric Apparatus ("order-of-magnitude" approximation only):

Based on assumptions 8, 9, and 10, this approximation is:

Per unit—$2,163.13×$\frac{0.30}{0.35}$×0.50 = $927.06
Nationally—$927.06×25,419 = $23,564,938

Dirt Damage on Other Portions ("order-of-magnitude" approximation only):

Based on assumptions 8, 9, and 10, this approximation is:

National total = $435,484,197×0.35×0.25 = $38,104,867

References

1. 1955 Survey of Diesel-Electric Motive Power in Railway Service and New 1954 Installations of Diesel-Electric Motive Power (book). Research Department, Simmons-Boardman Publishing Corporation, New York, N. Y., 1955.
2. 1956 Installations of Diesel-Electric Motive Power. Research Department, Simmons-Boardman Publishing Corporation, July 1957.
3. Seventieth Annual Report on Transport Statistics in the United States for the Year Ended December 31, 1956. Interstate Commerce Commission, Washington, D. C., p. 142.
4. *Ibid.*, p. 132.
5. *Ibid.*, p. 130.

Discussion

M. D. Henshaw (General Electric Company, Erie, Pa.): In presenting this paper, the author has performed a real service to the railroad industry, through this first display of the probable cost in money of dealing with the dirt problem on today's locomotive.

Since road units comprise two-thirds of all units in service, it is interesting to consider the author's estimates in terms of the annual cost for each road unit.

The interim cleaning cost of $5.48 when performed every other day (round trip basis) amounts to approximately $1,000 per road unit per year.

The monthly cleaning cost of $88.06 totals approximately $1,000 per road unit per year.

The total yearly cleaning cost of $2,000 per road unit is chargeable to engine house expense, and the author classes it as unavoidable in consideration of its relation to engine wear and to the ICC safety regulations.

It would seem that this considerable cleaning expense warrants still further mechanization of railroad cleaning facilities, and an intense effort on the part of the builders to minimize or to eliminate causes of oil and water leaks.

The author's figures for the annual cost of dirt exposure, stated in terms of life of parts and cost of shopping amount to approximately $2,500 per road unit for the diesel engine; $900 for the electric equipment; and $1,900 for "other"; or a grand total of $5,300 per unit per year; chargeable to repairs under ICC account M-311.

The author classes the $1,900 dirt exposure expense as unavoidable, leaving $3,400 per road unit per year as avoidable expense, to be saved by improved filtration of air for the engine and for the electric equipment.

The author's estimate of $139 million as the total national dirt cost for all yard and road units is approximately 15% of the total 1956 net income of $876 million for all roads.

The magnitude of these figures is a real challenge to the builders to do something about the problem on new designs and to work with the railroads on economic application of improved filters to existing units.

Wheel-Slip Detection in Railroad Braking

C. M. HINES
MEMBER AIEE

THE INVENTION of the compressed air brake by George Westinghouse in 1869 was followed rather promptly by investigations of wheel slip and wheel slide. This is reported in the famous Galton Westinghouse Tests of 1878 and 1879.[1] Prior to that time, wheel slip or slide was unknown on cars because the hand brakes used did not supply sufficient braking effort. Wheel slip frequently occurred during acceleration on locomotives, but this was very easy to detect due to the fact that the number of driving wheels was relatively few and they were usually connected in such a manner that all would slip in unison. Wheel slide in braking the locomotive was also not unknown since emergency stops were usually made by reversing the drivers of the locomotive and applying steam for compression braking, sometimes to the extent of spinning the drivers backwards.

The advent of power brakes made it impossible for the engineman to detect wheel slip or slide on the cars. While wheel-slip control during acceleration was actively studied[2,3] many years later with the advent of high tractive effort electric and diesel-electric locomotives, only wheel-slip control during braking will be treated in this paper.

At this point it is appropriate to define what we refer to as wheel slip. In the field of railway braking, wheel slip is any

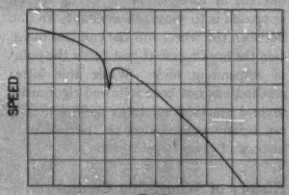


Fig. 1. Speed–time curve showing wheel-slip controller operation

Paper 58-303, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-ASME Railroad Conference, Cleveland, Ohio, April 9–10, 1958. Manuscript submitted December 9, 1957; made available for printing January 27, 1958.

C. M. Hines is with the Westinghouse Air Brake Company, Wilmerding, Pa.

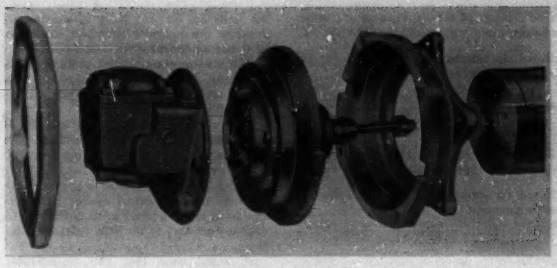


Fig. 2. Exploded view of mechanical inertia-type wheel-slip controller

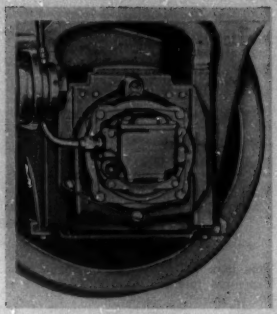


Fig. 3. Mechanical inertia-type wheel-slip controller installed on passenger car truck

departure from rolling contact between the wheel and rail in which the wheel continues to rotate but relative motion exists at the point of contact. In a wheel slide the same condition exists but the wheel no longer rotates. Slipping and sliding increase the stop distance for a train from any given speed due to the difference between the static and kinetic coefficients of friction. The static coefficient of friction for rolling contact on a dry rail is usually about 0.25 while the kinetic coefficient for sliding contact seldom is more than half that value. In addition to the undesirability of long stop distances, wheel sliding causes severe flat spots on the wheels and damages the surface contour of the rail.

The causes of wheel slip have been covered in many previous papers.[4,5] Briefly, wheel slips occur when the braking force between the brake shoe and the wheel exceeds the adhesion between the wheel and the rail. Poor adhesion is usually the result of foreign matter such as snow, dirt, or grease on the rail. Other causes are track irregularities, weight transfer, and unbalanced wheels.

The problem of wheel-slip control can be attacked by using either of two approaches. One is to improve the adhesion by any of several methods available.[6] This approach would compensate only for wheel slip caused by the introduction of foreign matter. The second approach is to sense the available friction between the wheel and the rail and maintain the braking effort just below the point of impending slide. While this latter method has not been followed rigorously, several practical solutions approaching it are briefly enumerated in the following history of protective braking.

History of Wheel-Slip Control

Early attempts at wheel-slip control were restricted to preventing wheel slide at the end of a stop. Such slides could usually be attributed to the variation in cast iron brake shoe friction with speed. The frictional force available from a cast iron shoe on a steel wheel varies inversely with speed from high to moderate speeds, increasing very rapidly as zero speed is approached. The first attempt at compensating for this variation in friction was made in 1910 when a test installation of an electric deceleration rate measuring device was applied to an interurban car. As the brake equipments in use at that time lacked the feature of being capable of release and reapplication without recharge, once the deceleration controller had caused a reduction in brake-cylinder pressure, this reduction in braking force would be in effect for the remainder of the stop. For this reason the controller was designed to reduce braking effort only during the latter part of the stop. While this device succeeded in increasing passenger comfort and decreased low-speed wheel slides, it could not recognize transient "bad rail" conditions at other speeds so wheel damage could still occur. For this reason no wheel-slip controllers of this type were ever offered as a standard device. However, in 1913 a pneumatic valve was introduced which served to reduce brake cylinder pressure gradually as train speed decreased. This was achieved by means of a variable orifice, the rate of reduction varying inversely with speed. The high-speed reducing valve made possible the application of high braking force at high speeds without danger of sliding as the train speed decreased. Many of these valves were installed in high-speed service, and while they had some of the shortcomings of the deceleration controllers mentioned previously, they provided a much needed means of obtaining a smooth stop without sliding and so played an important part in the history of protective braking.

In 1929 a pendulum-type deceleration measuring device was installed on streetcars (and subsequently on subway cars) in such a manner that excessive deceleration caused the pendulum to swing away from its neutral position and actuate a pilot valve or an electric contact which would then initiate a rapid reduction in brake cylinder pressure to a predetermined value. When the deceleration rate had decreased sufficiently, the pendulum then swung back, releasing the pilot valve or opening the contact and permitting the brakes to reapply. The streetcar brake equipment was of the straight-air-type, wherein the brake cylinder pressure was continuously maintained against leakage at the value called for by the brake valve. This feature made possible the reapplication of brake cylinder pressure referred to in the foregoing.

The pendulum was soon replaced with a large rectangular weight carried on rollers in an oil-filled housing. A change in velocity caused this weight to move on its rollers and actuate the pilot valve. Operation was otherwise identical to that of the pendulum system. As the deceleration sensing device was fastened to the carbody, it could respond only to decelerations of that body, and wheel sliding could still occur due to sudden shock, slack action, or changes in track conditions. The chief advantage was therefore the increase in passenger comfort.

In 1934 one of the early streamlined trains was equipped with a retardation control system. The retardation controller was of the rolling-weight type and was so located as to sense deceleration of the train as a unit. Its governing retardation rates were adjusted to provide smooth stops for the poorest rail condition that was likely to be encountered. For this reason the train could not be braked

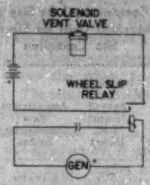


Fig. 4. D-c generator-type wheel-slip control schematic wiring diagram

to the full extent permissible with good rail conditions. The main difficulty encountered with the installation, however, was that it responded to the average deceleration of the train as a whole and could not recognize transient changes in friction occurring at individual wheel locations. These transient conditions are brought about by wheel bounce, track curves, body surge, and other contingencies contributing to unequal wheel loading. The result was that numerous wheel slips and slides could occur without the train deceleration exceeding its preset amount. In fact if several wheels of the train started to slide while a retardation controlled brake application was in effect, the reduction in retardation resulting from the sliding wheels would cause the controller to call for more braking effort to bring the retardation back to the predetermined value. The net result was that the braking force was thus increased, resulting in more wheel sliding.

The shortcomings of the retardation control system in high-speed passenger service brought about a system of proportioning braking effort with the friction available at various speeds by means of an empirical speed-brake cylinder pressure relationship.[4] This system, which is in current use, provides greatest brake force to the wheel at high speeds where the coefficient of friction between the wheel and cast iron shoe is lowest and reduces the brake force in steps as the speed declines. Smooth, minimum distance stops are thus obtained providing rail conditions remain normal for the duration of the stop. Transient bad rail or wheel loading variation conditions can still result in wheel slips and slides.

Passenger car brakes by 1937 had progressed to the point where the brake cylinder pressure was maintained against leakage by a local air reservoir. This made it possible to release and reapply locally any given brake cylinder as many times as was necessary during a stop. The way was thus paved for control of wheel slip at the wheel itself which had hitherto been considered impractical. Development therefore proceeded on devices which would be capable of sensing excessive angular deceleration of the wheel and initiating corrective action by means of instantaneous reduction in braking effort.

Wheel-slip controllers for protective braking were developed along both electrical and mechanical lines. Numerous types were developed in each category, but as the principles involved were similar, the following description can be separated into two parts: mechanical wheel-slip sensors and electrical wheel-slip sensors.

Mechanical Wheel-Slip Sensors

A mechanical wheel sensor consists of an inertia wheel which is driven by the railroad car axle through a spring device. The inertia wheel normally rotates at axle speed, the driving torque not being sufficient to deflect the driving spring. If the deceleration of the wheel is normal, the inertia wheel tries to continue rotating at a speed faster than axle speed but is restrained from doing so by the drive spring. If, however, the deceleration rate reaches about 12 miles per hour per second, a value which experience has shown to indicate an impending slide, the inertia of the flywheel is sufficient to overcome the restraining spring and the flywheel then rotates forward with respect to the axle

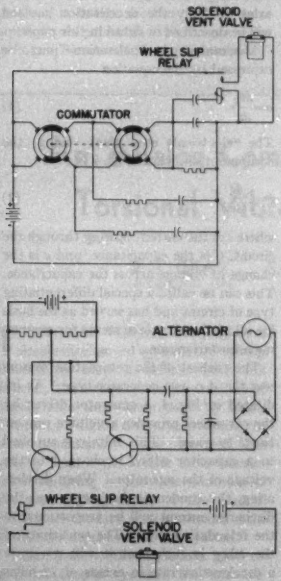


Fig. 5 (left). Commutator-type wheel-slip control schematic wiring diagram

Fig. 6 (below left). Electronic - type wheel-slip control schematic wiring diagram

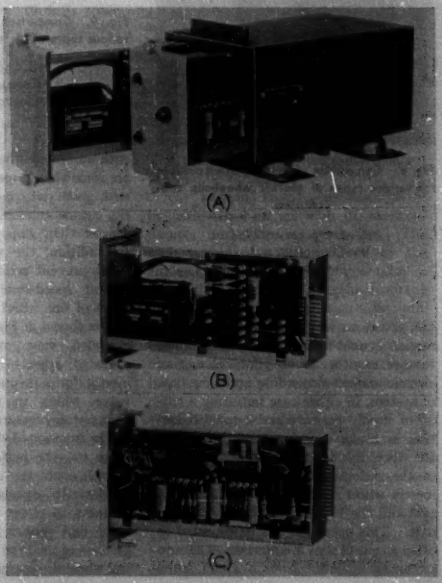


Fig. 7 (right). Control cabinet for electronic - type wheel-slip control

A—Partially disassembled view
B—Power supply chassis
C—Slip-sensing circuit chassis

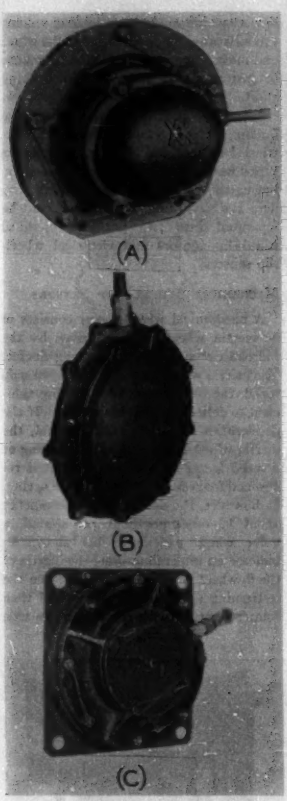


Fig. 8. Comparison of axle-mounted portions for various types of electric wheel-slip controllers

A—D-c generator type
B—Commutator type
C—Electronic type

and comes to rest against a stop. This motion is used by means of a cam surface to operate either an air pilot valve or an electric contact depending upon the type of system, in either case initiating a cut-off of supply pressure and a rapid reduction in brake cylinder pressure. When the wheel has resumed rolling as a result of the reduction in braking force, the inertia wheel returns to its neutral position and the brakes will reapply on that axle. This operation can be repeated as often as wheel slips are sensed. The stop distances achieved from a given speed

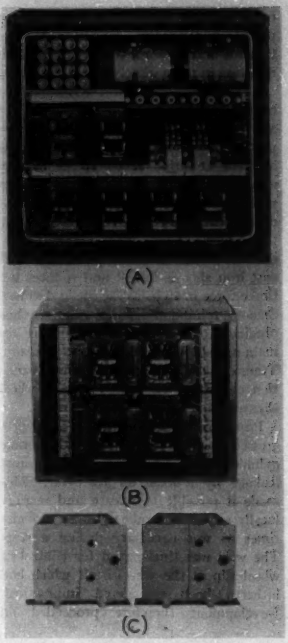


Fig. 9. Comparison of control cabinets for various types of electric wheel-slip controllers

A—D-c generator type
B—Commutator type
C—Electronic type

with numerous wheel-slip control releases are somewhat longer than those achieved with good rail conditions and no wheel-slip control releases, but are considerably shorter than stops where sliding occurs. In addition, the wheel-slip control releases prevent serious wheel damage. A typical speed-time curve showing the effect of one wheel slip controller operation is shown in Fig. 1. Fig. 2 shows an exploded view of a typical inertia-operated wheel-slip sensor where the inertia force is used to actuate a pilot valve which in turn controls brake cylinder air pressure.

The function of the cam lever and pilot valve may be performed with a pair of electric contacts. In this case excessive deceleration causes the normally open contacts to close, energizing a solenoid valve and venting brake cylinder. Regardless of the type of actuation, mechanical wheel slip sensors of this type are non-

Table I. Comparison of the Size and Weight of Axle-Mounted Units for Electric Wheel-Slip Controllers

Type of Controller	Projection from Face of Journal Box, Inches	Approximate Face Dimensions, Inches	Installation Weight, Pounds
Generator	7 1/8	13 diameter	87
Commutator	5 1/8	12 1/8 diameter	48
Electronic	3	9 by 9 1/8	17

directional by virtue of the symmetric design of the cylindrical cam surface.

Since the first mechanical wheel-slip controllers were installed, a great deal of progress has been made in improving the efficiency of the wheel-slip sensor. The first of these units installed for road service was over 14 inches in diameter and weighed about 130 pounds. The model in current use is less than 10 inches in diameter and weighs only 32 pounds. This device in a typical installation is shown in Fig. 3.

Electric Wheel-Slip Sensors

Wheel slip may be sensed electrically by two basic methods: measurement of deceleration or measurement of speed differential. In the latter method, the speeds of two axles are usually compared but methods have been evolved whereby the comparison is not limited to two axles.[2,3] Only the deceleration method will be described in detail in this paper.

Mechanically, deceleration may be expressed by the equation

$$a = \frac{dv}{dt} \tag{1}$$

The electrical equivalent using the equipment shown in Fig. 4 is

$$i = C\frac{de}{dt} \tag{2}$$

where i is the current flowing through the circuit, C is the capacitance, and e is the change in voltage across the capacitance. This can be called a special differentiating type of circuit and has served as the basis for all of the electric systems for measuring retardation rate.

The earliest of the retardation sensors was the d-c axle-generator-type. As indicated in Fig. 4, a generator driven by the car wheel provides a voltage proportional to speed. This voltage is supplied to a capacitor which is charged to the voltage of the generator. When decelerating, the condensor discharges and the discharge current will be proportional to the retardation rate. The sensitivity of the relay is such that it will recognize a deceleration rate in excess of 12 miles

per hour per second. Closing the contact of the wheel slip-relay energizes a magnet valve which vents brake cylinder pressure and thus allows the wheel to resume train speed. The relay then releases and the brake reapplies.

While the operation of this type of equipment was satisfactory, it was determined that the same results could be achieved more efficiently by using a commutator switch on the axle to provide the voltage proportional to speed. The schematic circuit diagram for one axle of this system is shown in Fig. 5. This scheme is identical to the preceding one except that the commutator and *R-C* circuits replace the generator in the basic circuit.

The advent of lightweight, high-speed trains created a need for a light, compact wheel-slip controller which could be mounted in spaces too restricted for the existing mechanical and electric controllers. An electronic wheel-slip controller was developed to fill this need. This controller utilizes the same basic differentiating circuit described previously, but uses the most modern components available to achieve maximum efficiency in a minimum space. A simplified circuit diagram for one axle is shown in Fig. 6. An impulse generator is used to provide a voltage proportional to speed. This voltage is rectified by means of a bridge rectifier and is then supplied to a capacitor as in the previous circuits. Because of the small amount of voltage generated, a 2-stage transistor amplifier is used to gain sufficient output to operate the wheel slip relay. The use of the impulse generator in place of a d-c generator or a commutator reduced the size of the wheel mounted portion and eliminated all slip rings and brushes.

The unitized design employed in the electronic wheel-slip controller can be seen in Fig. 7. The top view shows the two chassis partially removed from the cabinet, the center view shows the power supply chassis, and the bottom view the wheel-slip sensing chassis. While the circuits are designed for adaptation to printed circuit techniques, this type of construction is not at present in use.

Figs. 8 and 9 show the reduction in size which has been achieved during the years of development of wheel slip controllers. Fig. 8 shows the comparative sizes of the wheel mounted units with the oldest being at the top. Table I shows a comparison of the dimensions and weights for these units. Fig. 9 is a view of the controller cabinets required for a 4-axle car in the same order as the wheel units were shown. The principle of operation has remained basically unchanged, but changes in design and the availability of new devices has made possible a much more compact design. This is exemplified by the 5 by 12 inch space occupied by the two electronic cabinets shown at the bottom of Fig. 9 as compared to the 18 by 19 inch space occupied by the d-c generator controller cabinet shown at the top.

Conclusion

Wheel slip in braking has existed since the inception of power brakes. Much time and effort has been spent on the development of devices to minimize the effect of wheel slippage and it is impossible to recognize adequately the many individuals who have in the years past been associated with the work that has been accomplished. Continued effort has resulted in devices which have become more efficient through the use of the latest techniques available in the art. Wheel-slip control devices continue to help improve the performance of passenger car equipment with resulting economies to the railroads and increased comfort to passengers.

References

1. *Air Brake Tests.* Westinghouse Air Brake Company, Pittsburgh, Pa., 1904.

2. Locomotive Wheel-Slip and Wheel-Lock Protection, R. M. Smith. *AIEE Transactions*, vol. 39, pt. II, 1950, pp. 1154–57.

3. A New Wheel-Slip Protective Scheme, R. M. Smith. *Ibid.*, vol. 71, pt. II, 1952, pp. 51–53.

4. Electropneumatic Brakes for High-Speed Trains with Particular Reference to Their Electrical Features, Joseph C. McCune. *Ibid.*, vol. 61, March 1942, pp. 137–41.

5. Development and Testing of Brakes for High-Speed Railroad Equipment, C. E. Tack. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 73, no. 2, Feb. 1951, pp. 167–82.

6. Investigation of Causes of Low Wheel-to-Rail Adhesion and Possible Methods of Improving It, F. G. Fisher, R. K. Allen. *Ibid.*, vol. 79, no. 2, Feb. 1957, pp. 377–88.

An Analog Computer Study of a Torsional Vibrations Problem

R. T. GRAY
NONMEMBER AIEE

S. W. McELHENNY
MEMBER AIEE

TORSIONAL VIBRATION studies of simple mechanical systems are usually handled by methods which, while not difficult, often require the expenditure of considerable time when done by hand. The use of a digital computer shortens the time considerably. One method frequently used is the Holzer table calculations[1] resulting in the determination of the natural frequencies of the system. Another method is a matrix solution of the differential equations of the system.

While these methods produce good results for simpler systems, the introduction of nonlinear driving functions, discontinuities in the form of gear backlash, and system components, such as hydraulically-connected vibration dampers, often require simplifying assumptions which affect the validity of the results. The proposed method overcomes some of these difficulties and predicts the actual operation of the system more accurately. Using the analog computer, nonlinearities and discontinuities in the system function, as well as in the driving function, can be accounted for. This amounts to the particular (steady-state) solution of the differential equations of the system. The complementary (transient) solution can also be obtained.

The system studied includes a 6-cylinder diesel engine with its torsional damper and two d-c generators. The main generator is direct driven by the engine, while the auxiliary generator is driven through a set of spur gears from the main generator. A simplified schematic diagram of the system is shown in Fig. 1. For simplicity, the expressions "mass moment of intertia" and "mass" have, in many instances, been used interchange-

Paper 58-502, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-ASME Railroad Conference, Cleveland, Ohio, April 9–10, 1958. Manuscript submitted December 30, 1957; made available for printing February 17, 1958.

R. T. Gray and S. W. McElhenny are with the General Electric Company, Erie, Pa.

The authors acknowledge with appreciation the assistance rendered by R. N. Nichael in devising analog computer circuits, particularly for the engine distributor.

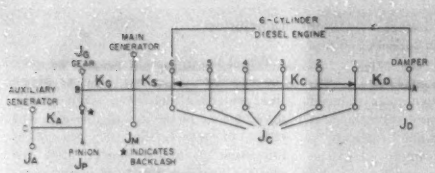


Fig. 1. Schematic of equivalent system. All gear ratios have been taken into account on J and K. Engine-firing order: 1, 5, 3, 6, 2, 4. Bore = 9 inches, stroke = 10.5 inches

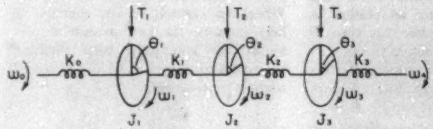


Fig. 2. Schematic of a typical system

T_1, T_2, and T_3 are the externally applied driving or load torques
J_1, J_2, and J_3 are the moments of inertia of the lumped masses
ω_1, ω_2, and ω_3 are the angular velocities of the masses
θ_1, θ_2, and θ_3 are the angular displacements of the masses with respect to an arbitrary datum
K_1, K_2, and K_3 are the torsional spring constants of the connecting links between the various lumped masses

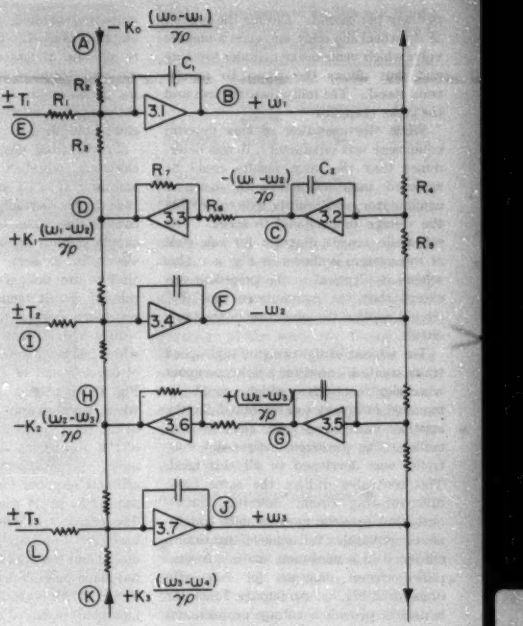


Fig. 3 (right). Computer simulation of typical system

ably throughout this paper. Both the masses and spring constants are shown in terms of the main shaft angular velocity.[2] The system is broken into discrete lumps and the standard practice for analyzing reciprocating engines is used. The rotating and reciprocating parts are lumped to form average equivalent moments of inertia. The actual system is distributed, of course, and not lumped.[1]

Assumptions

For purposes of analysis, the following assumptions have been made:

1. A distributed mass system can be represented by discrete lumped masses which are elastically connected.
2. Friction and damping can be neglected, except in the case of the engine damper.
3. Engine torque magnitude at any crank angle varies lineally as a function of fuel.
4. The compression portion of the engine-torque-versus-crank-angle characteristic remains the same regardless of speed or fuel.
5. Torsional spring constants of the gear teeth can be neglected.
6. Moments of inertia of the pinions can be lumped with adjacent moments.
7. The moment of inertia of the torsional vibration damper housing can be neglected.
8. The vibration recorder has no effect on the system.

Analyses

Basic Spring-Mass Simulation

A lumped-mass torsional spring analogy of a simple mechanical system is shown in Fig. 2, and is used as the basis for the analog computer simulation. The application of Newton's second law enables the torque equations to be written in terms of computer time; see the Appendix. From equation 71 in the Appendix it can be stated that:

$$\omega_1 = \frac{1}{\gamma\rho J_1}\left[T_1 + K_0\frac{1}{\gamma\rho}(\omega_0-\omega_1) - K_1\frac{1}{\gamma\rho}\times(\omega_1-\omega_2)\right] \quad (1)$$

$$\omega_2 = \frac{1}{\gamma\rho J_2}\left[T_2 + K_1\frac{1}{\gamma\rho}(\omega_1-\omega_2) - K_2\frac{1}{\gamma\rho}(\omega_2-\omega_3)\right] \quad (2)$$

$$\omega_3 = \frac{1}{\gamma\rho J_3}\left[T_3 + K_3\frac{1}{\gamma\rho}(\omega_2-\omega_3) - K_3\frac{1}{\gamma\rho}(\omega_3-\omega_4)\right] \quad (3)$$

The analog simulation for this system is shown in Fig. 3. If the system is composed of more masses, they may be simulated similarly and connected in the proper locations in a ladder fashion. It should be noted that the signs of ω alternate throughout Fig. 3, but this results in a considerable saving of operational amplifiers.

A through L in Fig. 3 indicate magnitude scale factors. Torques T_1 through T_3 may be either sign, depending upon whether they are driving or load torques.

Amplifiers 3.1, 3.4, and 3.7 sum, integrate and divide by the appropriate J's to produce the ω's.

Amplifiers 3.3 and 3.6 multiply $\Delta\theta$'s by the appropriate torsional spring constants K's to produce the spring torque applied to the various masses.

Thus, solving for the values of the passive components from equation 1 for amplifier 3.1 in Fig. 3; see Appendix for basic computer equation:

$$\frac{1}{\rho R_1 C_1} = \frac{1}{J_1\gamma\rho}\times\frac{E}{B} \quad \text{or} \quad (4)$$

$$R_1C_1 = \frac{B}{E}\times J_1\gamma \quad \text{and} \quad (5)$$

$$R_2C_2 = \frac{B}{A}\times J_1\gamma \quad (6)$$

$$R_3C_3 = \frac{B}{D}\times J_1\gamma \quad (7)$$

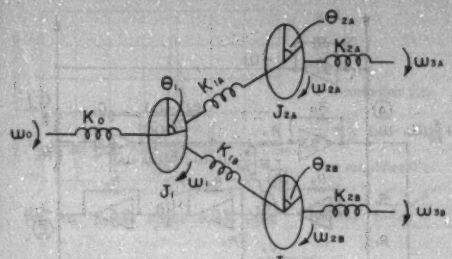


Fig. 4. Gear branching

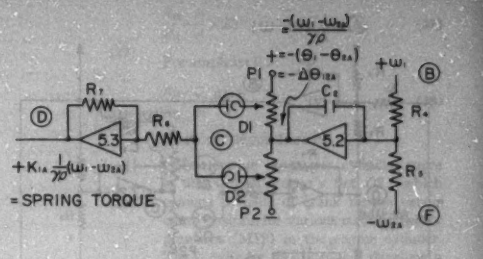


Fig. 5. Backlash simulation

Similarly for amplifier 3.2

$$\frac{1}{\rho R_4 C_3}=\frac{1}{\gamma\rho}\times\frac{B}{C} \tag{8}$$

or

$$R_4C_3=\frac{C}{B}\gamma \tag{9}$$

and

$$R_5C_3=\frac{C}{F}\gamma \tag{10}$$

and for amplifier 3.3, which only multiplies

$$\frac{R_7}{R_6}=\frac{C}{D}\times K_1 \tag{11}$$

Additional passive component values may be solved for similarly.

Simulation of Gear Branching with Blacklash

A preliminary investigation showed that, with reasonable contact area between meshing gear teeth, the elastic deformation of the gear teeth is negligibly small with respect to the associated pinion and gear shaft angular displacements.[3] That is, the torsional spring constant of the gear teeth is infinitely large as compared to the torsional spring constants of the associated shafting for this particular system. Consequently, the gear teeth spring constants are neglected. Also, the moments of inertia of the pinions are lumped with the appropriate adjacent moment of inertia. The torsional analogy for a general system of a gear and two pinions is shown in Fig. 4.

Referring to Fig. 4, J_1 is the moment of inertia of the gear, while J_{2A} and J_{2B} are the moments of inertia of each of the pinions lumped with its adjacent moment of inertia. K_{1A} and K_{2B} are the torsional spring constants of the shafts between each of the pinions and their adjacent moments of inertia, while K_0 is the torsional spring constant of the shaft driving the gear. K_{1A} and K_{2B} are the torsional spring constants of the links going away from the moments of inertia represented by J_{2A} and J_{2B}. ω_1, ω_{2A} and ω_{2B} are the angular velocities of the respective lumped masses while θ_1, θ_{2A} and θ_{2B} are the respective angular displacements with respect to an arbitrary datum.

The angular velocities of the pinion branches usually differ from the angular velocity of the main branch of the system by a gear ratio. Because the main branch of the system has the greatest number of lumped masses, the moments of inertia and the torsional spring constants are referred back to the main branch.

During the period that the teeth of a pinion are disengaged from the gear teeth, the system can be thought of as two smaller independent systems. Thus, the spring K_{1A} (Fig. 4) can be considered nonexistent when the pinion teeth of branch $2A$ are disengaged with the gear teeth such that branch $2A$ and the remainder of the system act independently during backlash. The same analysis is true for branch $2B$. A circuit for the simulation of backlash is shown in Fig. 5, while a sketch of the meshing teeth is shown in Fig. 6.

Referring to Fig. 6, if the gear and pinion are in the position shown, the pinion must move a distance of $\pm\Delta\theta/2$ in order for the teeth to engage. Until this angular distance is traversed, no torque is transmitted in either direction. This distance is determined from tooth clearances, and potentiometers $P1$ and $P2$, in Fig. 5 are set to bias diodes $D1$ and $D2$ so that the output of amplifier 5.2 must be equal to either $\pm\Delta\theta/2$ (as shown in Fig. 6), before any signal is applied to amplifier 5.3. Thus, no torque is transferred from the main shaft to the branch (or vice versa) until the relative angle between the pinion and gear simulations is greater than $\Delta\theta/2$.

The analog simulation of the torsional system shown in Fig. 4 is shown in Fig. 7. Here A through G indicate magnitude scale factors. Amplifiers 7.1, 7.4, and 7.7 sum, integrate and divide by the appropriate J to produce the ω's. Amplifiers 7.2 and 7.5 subtract and integrate the adjacent ω's to produce the $\Delta\theta$'s. Amplifiers 7.3 and 7.6 multiply the $\Delta\theta$'s by the appropriate torsional spring constants K's to produce the spring torques applied to the various masses.

The basic equations are (in terms of computer time):

$$\omega_1=\frac{1}{\gamma\rho J_1}\left[K_0\frac{1}{\gamma\rho}(\omega_0-\omega_1)-K_{1A}\frac{1}{\gamma\rho}\times(\omega_1-\omega_{1A})-K_{1B}\frac{1}{\gamma\rho}(\omega_1-\omega_{1B})\right] \tag{12}$$

$$\omega_{1A}=\frac{1}{\gamma\rho J_{1A}}\left[K_{1A}\frac{1}{\gamma\rho}(\omega_1-\omega_{1A})-K_{1A}\frac{1}{\gamma\rho}(\omega_{1A}-\omega_{1A})\right] \tag{13}$$

$$\omega_{1B}=\frac{1}{\gamma\rho J_{1B}}\left[K_{1B}\frac{1}{\gamma\rho}(\omega_1-\omega_{1B})-K_{1B}\frac{1}{\gamma\rho}(\omega_{1B}-\omega_{1B})\right] \tag{14}$$

Solving for the values of the passive components from equation 12 and Fig. 7 for amplifier 7.1:

$$\frac{1}{\rho R_1 C_1}=\frac{1}{\gamma\rho J_1}\times\frac{A}{B} \tag{15}$$

or

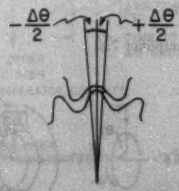


Fig. 6. Meshing teeth, showing backlash

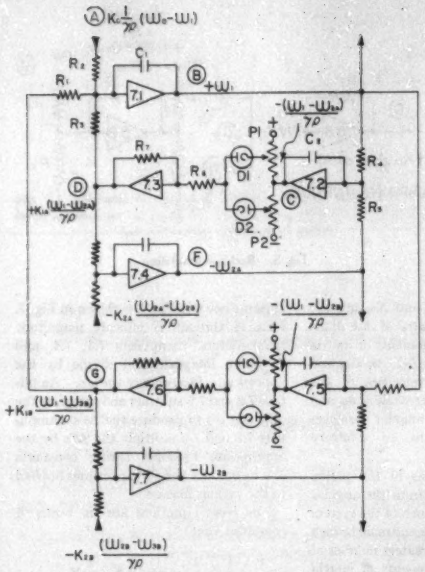


Fig. 7. Computer simulation of gear branching with backlash

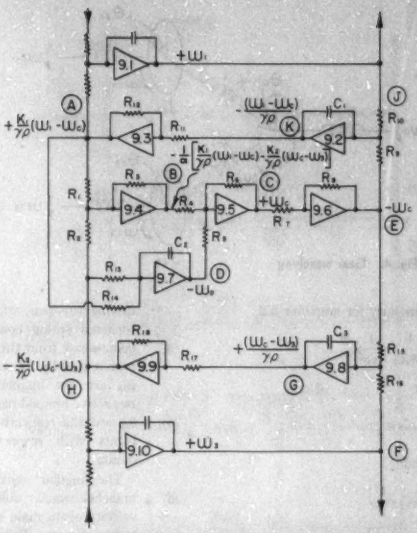


Fig. 9. Computer simulation of torsional vibration damper

$$R_2C_1=\frac{B}{A}\times J_1\gamma \tag{16}$$

and

$$R_3C_1=\frac{B}{D}\times J_1\gamma \tag{17}$$

$$R_1C_1=\frac{B}{G}\times J_1\gamma \tag{18}$$

Similarly for amplifier 7.2:

$$\frac{1}{\rho R_4C_2}=\frac{1}{\gamma\rho}\times\frac{B}{C} \tag{19}$$

or

$$R_4C_2=\frac{C}{B}\gamma \tag{20}$$

and

$$R_5C_2=\frac{C}{F}\gamma \tag{21}$$

and for amplifier 7.3:

$$\frac{R_7}{R_6}=\frac{C}{D}\times K_{1A} \tag{22}$$

additional passive component values may be solved for similarly.

Additional pinions driven from the gear may be handled similarly. That is, the additional branches may be connected to amplifier 7.1 as are the A and B branches.

Simulation of Torsional Vibration Damper

The front end of this particular diesel engine includes a torsional vibration damper. It is a viscous-type damper composed of a can shaped like a hollow ring, concentric with and driven by the crank shaft. The can encloses a doughnut-shaped ring floating in oil. This is shown schematically in Fig. 8.

The damper operation is similar to a low pass filter in that the higher frequency torsional vibrations are attenuated by it. The mass of the can for this particular damper is very small and consequently is neglected. The equation for the damper then is simply that of a single time constant circuit having a break frequency of α/J_D radians per second where α is the damping coefficient in force-length-time units and J_D is the moment of inertia of the doughnut. J_C is the moment of inertia of the can, J_1 and J_3 are the moments of inertia of the preceding and succeeding masses having angular velocities of ω_1 and ω_3. ω_C and ω_D are the angular velocities of the can and doughnut.

T_1 and T_3 are externally applied torques. K_0, K_1, K_2, and K_3 are the torsional spring constants of the links in the system.

By applying Newton's second law, the spring torque equations may be written in terms of computer time; see the Appendix. Rearranging equations 75, 76, and 77 in the Appendix gives:

$$\omega_C=\frac{1}{\alpha}\left[K_1\frac{1}{\gamma\rho}(\omega_1-\omega_C)-K_2\frac{1}{\gamma\rho}(\omega_C-\omega_3)\right]+\frac{1}{J_D\gamma\rho}\left[K_1\frac{1}{\gamma\rho}(\omega_1-\omega_C)-K_2\frac{1}{\gamma\rho}(\omega_C-\omega_3)\right] \tag{23}$$

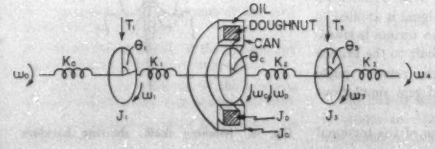


Fig. 8. Torsional vibration damper

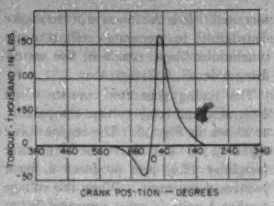


Fig. 10. Engine torque versus crank

The analog simulation for this system is shown in Fig. 9. A through K indicate magnitude scale factors. Amplifier 9.4 sums the two spring torques and divides by α to produce $1/\alpha[K_1 1/\gamma\rho(\omega_1-\omega_c)-K_2 1/\gamma\rho(\omega_c-\omega_3)]$. Amplifier 9.7 sums the same two spring torques, divides by J_D and integrates to produce ω_3. These two quantities are summed in amplifier 9.5 to produce ω_c. Amplifier 9.6 is a sign changer. Amplifiers 9.2 and 9.8 sum and integrate the ω's to produce $\Delta\theta$'s. These are then multiplied by their appropriate torsional spring constants in amplifiers 9.3 and 9.9 to produce the two spring torques which operate on the can.

Solving for the values of the passive components using equation 23 and Fig. 9 for amplifier 9.4:

$$\frac{R_3}{R_1}=\frac{1}{\alpha}\times\frac{A}{B} \tag{24}$$

and

$$\frac{R_2}{R_3}=\frac{1}{\alpha}\times\frac{H}{B} \tag{25}$$

For amplifier 9.5:

$$\frac{R_5}{R_4}=\frac{B}{C} \tag{26}$$

and

$$\frac{R_4}{R_5}=\frac{D}{C} \tag{27}$$

For amplifier 9.6:

$$\frac{R_7}{R_7}=\frac{C}{E} \tag{28}$$

For amplifier 9.7:

$$\frac{1}{\rho R_{14}C_3}=\frac{1}{\gamma\rho J_D}\times\frac{A}{D} \tag{29}$$

or

$$R_{14}C_3=\frac{D}{A}\times J_D\gamma \tag{30}$$

and

$$R_{13}C_3=\frac{D}{H}\times J_D\gamma \tag{31}$$

For amplifier 9.2:

$$\frac{1}{\rho R_9C_1}=\frac{1}{\gamma\rho}\times\frac{E}{K} \tag{32}$$

or

$$R_9C_1=\frac{K}{E}\times\gamma \tag{33}$$

and

$$R_{10}C_1=\frac{K}{J}\times\gamma \tag{34}$$

For amplifier 9.8:

$$R_{15}C_2=\frac{G}{E}\times\gamma \tag{35}$$

and

$$R_{16}C_2=\frac{G}{F}\times\gamma \tag{36}$$

For amplifier 9.3:

$$\frac{R_{12}}{R_{11}}=\frac{K}{A}\times K_1 \tag{37}$$

For amplifier 9.9:

$$\frac{R_{18}}{R_{17}}=\frac{G}{H}\times K_2 \tag{38}$$

Engine Torque Inputs

The torque input per cylinder curve is taken from Porter's Set $Q2$,[4] which consists of plots of crank torque versus crank position for various mean indicator pressures (MIP) in the engine cylinder. The 6-cylinder diesel engine develops a maximum of 197 MIP, but Porter's curves go only as high as 150 MIP. Consequently Porter's 150 MIP curve is extrapolated to 197 MIP for the full load power portion (input torque). The compression portion of the cycle (loading torque) is taken directly from Porter's curves. During computer operation, the ordinates of the power portion of the 197 MIP curve are continuously multiplied by a fuel factor from the governor. This factor, which is continuously determined by the governor (see succeeding section on "Engine Governor Simulation"), keeps the system at any desired constant speed by making the output power of the engine equal to the sum of that power required by the various loads. This results in an approximation of the actual input torque per cylinder because with increasing cylinder MIP the peak of the actual torque pulse occurs later in the cycle,[4] however, the method used was the most expedient compromise to obtain changes in the various loadings on the simulated engine and still be a practical simulation of the torque pulse per cylinder. The curve of torque in inch-pounds

Fig. 11 (below). Computer circuit for two engine torque pulses

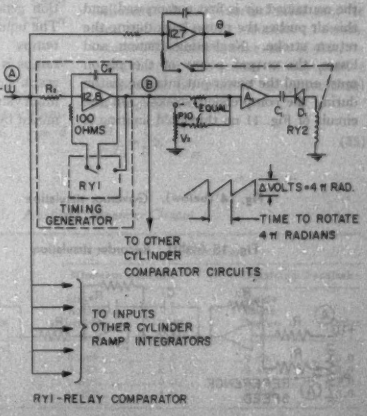
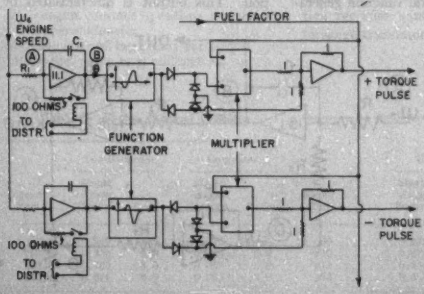


Fig. 12 (right). Schematic of engine distributor



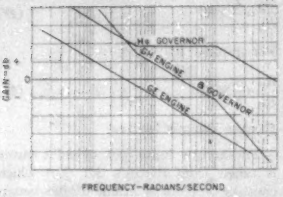


Fig. 13. Frequency attenuation characteristic

versus crank position in degrees is shown by Fig. 10.

The torque pulses are generated with six electronic function generators one pulse function for each cylinder in the engine. The 1, 5, 3, 6, 2, 4 firing order of the engine is simulated by the distributor (see succeeding section on "Engine Distributor"), such that the six cylinders are fired at equal intervals within two revolutions of the engine. For each cylinder, the distributor triggers off the integration of a constant voltage proportional to engine speed. This results in an integrator output proportional to crank degree position. The integrator output feeds the X input of the function generator for a particular cylinder and the function generator output is fed through a diode network, a multiplier and a summer to the equivalent inertia of that cylinder. Two of the six function generators with their associated circuits are shown in Fig. 11.

The diodes shown in Fig. 11 (on the outputs of the function generators) are to prevent the compression portion of the curve from being multiplied by the fuel factor from the governor; this portion of the curve remains the same regardless of the power output of the engine. Actually with no fuel entering the cylinder, the the contained air is first compressed, and this air pushes the piston back during the return stroke. Neglecting friction and losses, the output power of the piston must equal the power put into the piston during the compression stroke. On the circuit of Fig. 11 no threshold limiting is used on the output power portion of the simulated engine. The only error resulting from this is during the deceleration of the engine when the engine output power is less than the input power required for compression. This does no harm because readings are taken only for steady-state speeds and at small incremental increases of speed. In order to take accurate transient readings (both acceleration and deceleration), the input signal from the simulated governor would have to be limited to a certain minimum. The approximate friction torque was calculated and found to be an insignificant portion of the net output of the engine, and consequently, it is neglected.

Fig. 11 shows that one of the outputs is positive and the next is negative. Because of the alternate signs of the computer loops which simulate the system (as explained in the first section), it is necessary to have alternating positive and negative input signals.

The equation represented by amplifier 9.1 is (in terms of computer time):

$$\theta = \frac{\omega}{\gamma \rho} \tag{39}$$

Therefore:

$$\frac{1}{\rho R_1 C_1} = \frac{A}{B} \times \frac{1}{\gamma \rho} \tag{40}$$

or

$$R_1 C_1 = \frac{B}{A} \gamma \tag{41}$$

where A and B are the magnitude scale factors.

Engine Distributor

The engine torque inputs required for the simulation are supplied by six function generators (one for each cylinder). The inputs to these generators are linear ramps. The duration of each of the ramps is equal to two revolutions of the crank shaft or 4 π radians. The maximum value of the ramps (Δ volts) is determined by the particular function generators used. It is the function of the engine distributor to generate and properly synchronize these ramps in the correct firing order (1, 5, 3, 6, 2, 4).

The timing generator consists of an integrator and relay voltage comparator, as shown in Fig. 12. The engine speed signal ω is fed into the integrator (amplifier 12.8) and produces a linear positive going ramp. When the voltage reaches Δ volts (a value corresponding to 4 π radians for a 4-stroke cycle engine), the relay voltage comparator closes, discharges the integrator capacitor and resets integrating amplifier 12.8 output to zero. As the relay voltage comparator has a differential of only 50 microvolts, its contacts open almost instantaneously and the next ramp starts immediately.

The integrator RC product is selected so that with a given ω the amplifier will integrate to Δ volts in exactly 4 π radians. The 100-ohm resistor in the relay contact circuit limits the peak current during the capacitor discharge.

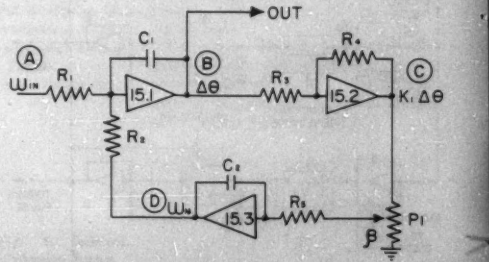
To supply the inputs to the function generators, six integrators identical to that in the timing generator are used. The inputs to these integrators are fed from the same ω as the timing generator. Six voltage comparator circuits are used to synchronize the starting of these ramps in the proper firing order. Only one of these circuits is shown in Fig. 12.

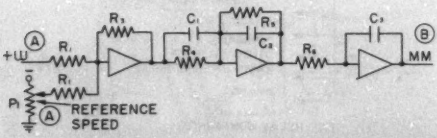
The ramp from the timing generator represents 4 π radians (two revolutions of the crankshaft) and has a maximum value of Δ volts. There are six cylinders and each one fires once in two revolutions; therefore, the starting of the individual cylinder ramps is spaced to start at intervals of $\Delta/6$ volts.

The reference voltage V_3 from potentiometer P-10 is set at $-\Delta/12$ volts. As there is no feedback on amplifier $A1$ (an unstabilized amplifier) it is driven to positive as V_1 increases above $\Delta/12$ volts. At this point, amplifier $A1$ output will change from positive to negative saturation. This output is differentiated by

Fig. 14 (below). Governor simulation

Fig. 15 (right). Recorder simulation





capacitor $C1$ and the resulting pulse momentarily closes relay $RY2$ which shorts the feedback capacitor on amplifier 12.7 and resets the amplifier output to zero. Here again, the 100-ohm resistor is for current limiting. The diode $D1$ in the relay circuit prevents $RY2$ from pulsing when the main timing generator resets, causing the output from Amplifier $A1$ to change from negative to positive saturation.

The circuits for the other five cylinders are similar except for the reference voltage $V2$ which is set at

$$-\left(\frac{2\Delta}{6}-\frac{\Delta}{12}\right),\ -\left(\frac{3\Delta}{6}-\frac{\Delta}{12}\right),\ -\left(\frac{4\Delta}{6}-\frac{\Delta}{12}\right),$$
$$-\left(\frac{5\Delta}{6}-\frac{\Delta}{12}\right),\ -\left(\frac{6\Delta}{6}-\frac{\Delta}{12}\right)$$

respectively.

The ramps are thus held in time with each other and with the timing generator regardless of the input ω. Amplifier 12.7, Fig. 12 is the same as amplifier 11.1, Fig. 11.

The equation represented by amplifier 12.8 is (in terms of computer time):

$$\theta=\frac{\omega}{\gamma p} \qquad \text{Therefore} \tag{42}$$

$$\frac{1}{\rho R_3 C_3}=\frac{A}{B}\times\frac{1}{\gamma p} \qquad \text{or} \tag{43}$$

$$R_3C_3=\frac{B}{A}\times\gamma \tag{44}$$

Engine Governor Simulation

Speed regulation of the system is accomplished by an engine governor simulation. The governor is an isochronous integrating type having a numerator and a denominator time constant to provide stability.

The approximate transfer function for the engine is of the form:

$$G_E=\frac{\Delta\omega}{\Delta MM}=\frac{K_E}{p} \tag{45}$$

where

K_E = engine constant in radians per second per millimeter
ω = engine speed in radians per second
MM = fuel pump rack millimeters

This characteristic is plotted in Fig. 13.

The governor and the engine are both integrating devices and will cause the open loop gain characteristic to attenuate through unity at a 40 db (decibels) per decade rate indicating −180-degree phase shift or 0 degree phase margin. Consequently, the governor is made to have the transfer function of equation 46, shown plotted in Fig. 13. This achieves more phase margin at unity gain to provide for stable operation.

$$H_G=\frac{\Delta MM}{\Delta\omega}=\frac{K_G(1+T_1p)}{p(1+T_2p)} \tag{46}$$

where K_G = governor constant in millimeters per radian. The open loop function then becomes:

$$G_EH_E=\frac{\Delta\omega}{\Delta\omega}=\frac{K_EK_G(1+T_1p)}{p^2(1+T_2p)} \tag{47}$$

which is also shown plotted in Fig. 13.

The computer simulation for the governor only is shown in Fig. 14.

In general, equation 46 becomes (in terms of computer time):

$$H_G=\frac{\mu(1+T_1\gamma p)}{\gamma p(1+T_2\gamma p)} \tag{48}$$

Solving for the values of the passive components using equation 48 and the input and feedback impedances of the amplifiers in Fig. 14:

$$\frac{R_2R_3}{R_1R_4R_5C_3}(1+\gamma R_4C_4)\frac{1}{\rho(1+\rho R_3C_3)}=\frac{\mu(1+T_1\gamma p)}{\gamma p(1+T_2\gamma p)} \tag{49}$$

$$R_4C_4=T_1\gamma \tag{50}$$

$$R_3C_3=T_2\gamma \tag{51}$$

$$\frac{R_2R_3}{R_1R_4R_5R_3}=\frac{\mu}{\gamma}\times\frac{A}{B} \tag{52}$$

Potentiometer $P1$ provides the speed reference for the governor. It is set for the desired value of ω.

Recorder Simulation

The test data were taken from the actual system using a Cox-Type 4 vibration recorder equipped with a Type 4 torsional translator.

The recorder[3] consists mainly of a very light pulley driven by a nonelastic belt from the shaft under investigation. An inertial mass is loosely coupled elastically to the light pulley such that its angular velocity remains relatively constant because of its inertia.

The relative motion between the mass and pulley is proportional to rotational fluctuations and this motion is recorded on a moving chart by a stylus. These displacements can then be converted into angular measurements describing the instantaneous relative positions of the shaft under investigation.

In order to get a comparison between computer results and test data, it was necessary to simulate the recorder. The computer simulation is shown in Fig. 15. Output of amplifier 15.1 is the difference in angular position $\Delta\theta$ between the integral of ω_{IN} (the angular velocity of the mass under investigation) and the integral of ω_M (the angular velocity of the inertial mass). Expressed as an equation:

$$\Delta\theta=\frac{\omega_{IN}-\omega_M}{p} \tag{53}$$

Amplifier 15.2 multiplies $\Delta\theta$ by the torsional spring constant K_1 of the spring driving the inertial mass. Amplifier 15.3 integrates the spring torque ($K_1\Delta\theta$) and divides by the moment of inertia of the inertial mass (J_M) to produce the angular velocity of the inertial mass or:

$$\omega_M=\frac{K_1\Delta\theta}{\rho J_M} \tag{54}$$

Fig. 15 shows that there is no spring torque ($K_1\Delta\theta$) coupled back into the mass under investigation. This is a reasonable assumption as the torsiograph has no measurable effect upon the system. Different values of inertial mass may be used with the actual recorder. Potentiometer $P1$ is used to take care of this.

Solving for the values of the passive components (in terms of computer time) for amplifier 15.3:

$$\frac{\beta}{\rho R_3C_3}=\frac{C}{D}\times\frac{1}{\gamma\rho J_M} \tag{55}$$

Table I. Engine Damper Output

Speed, Rpm	Cox Recorder: Single Angle Degree	Cox Recorder: Order	Computer: Single Angle Degree	Computer: Order
1,000	0.249	4.5	0.230	4.5
750	0.200	6.0	0.115	6.0
612	0.103	6.0	0.063	6.0
460	0.071	9.0	0.029	9.0

Table II. Auxiliary Generator Output

Speed, Rpm	Cox Recorder: Single Angle Degree	Cox Recorder: Order	Computer: Minimum Backlash: Single Angle Degree	Computer: Minimum Backlash: Order	Computer: Maximum Backlash: Single Angle Degree	Computer: Maximum Backlash: Order
1,000	0.475	2.5–3.0	0.289	5.0	0.635	3.0
750	0.475	3.0	0.252	5.0	0.437	3.0
612	0.350	3.0	0.313	5.0	0.403	3.0
460	0.302	3.0	0.252	4.5	0.376	3.0

Table III. Generator Gear Output

Speed, Rpm	Cox Recorder: Single Angle Degree	Cox Recorder: Order	Computer: Single Angle Degree	Computer: Order
1,000	0.065	3.0	0.086	3.0
750	0.065	3.0	0.086	3.0
612	0.065	3.0	0.172	3.0
460	0.065	3.0	0.172	3.0

or

$$\frac{R_3C_3}{\beta}=\frac{D}{C}\times J_M\gamma \tag{56}$$

For amplifier 15.1:

$$\frac{1}{\rho R_1C_1}=\frac{A}{B}\times\frac{1}{\gamma\rho} \tag{57}$$

or

$$R_1C_1=\frac{B}{A}\times\gamma \tag{58}$$

and

$$R_2C_1=\frac{B}{D}\times\gamma \tag{59}$$

For amplifier 15.2:

$$\frac{R_4}{R_3}=\frac{B}{C}\times K_1 \tag{60}$$

Results

During the simulation, the actual system from which test data were obtained was available. A Cox recorder[6] was used to take recordings of the damper output, main generator shaft output, and auxiliary generator output (points A, B, and C in Fig. 1). The Cox recorder was simulated, and readings were also taken at the same locations of the simulation.

Comparisons of test and computer data are given in Tables I, II, and III. Single angle degrees is defined as one half the peak-to-peak relative angular displacement which one would see using a stroboscope. This is a magnitude measurement. Order is defined as the number of vibrations per revolution of the engine shaft and is a measure of vibration frequency. There appears to be good correspondence between test and computer data of both magnitude and frequency.

An interesting result is evident in Table II. Both the order and amplitude of vibration are affected by the amount of backlash between gear teeth. Only qualitative results have been obtained; but they indicate that, as backlash is increased, the vibration amplitude increases and the vibration frequency changes up to a point. (After this paper was written the authors' attention was called to the fact that similar conclusions were reached by Tuplin.[4]) When this point is reached, the frequency appears to hold constant but a further increase in amplitude results as the backlash is further increased. Quantitative relationships for this have not yet been established.

Conclusions

Results indicate that this method produces better results for a complex system than the Holzer calculation because nonlinearities and discontinuities can be included.

Further studies should be made to find more accurate methods of determining the mathematical expressions which represents the physical system components.

Appendix

Considering the typical system shown in Fig. 2, Newton's second law gives:

$$T_1+K_1(\theta_1-\theta_2)-K_2(\theta_2-\theta_3)=J_2\rho\omega_2$$

or

$$T_1+K_1\frac{1}{p}(\omega_1-\omega_2)-K_2\frac{1}{p}(\omega_2-\omega_3)=J_2\rho\omega_2 \tag{61}$$

where

$$p(\ )=\frac{d(\ )}{dt} \tag{62}$$

and

$$\frac{1}{p}(\ )=\int(\ )dt \tag{63}$$

Rearranging equation 61 and solving ω_2

$$\omega_2=\frac{1}{\rho J_2}\left[T_2+K_1\frac{1}{p}(\omega_1-\omega_2)-K_2\frac{1}{p}(\omega_2-\omega_3)\right] \tag{64}$$

Likewise:

$$\omega_1=\frac{1}{\rho J_1}\left[T_1+K_2\frac{1}{p}(\omega_2-\omega_1)-K_1\frac{1}{p}(\omega_1-\omega_2)\right] \tag{65}$$

and

$$\omega_3=\frac{1}{PJ_3}\left[T_3+K_2\frac{1}{p}(\omega_2-\omega_3)-K_3\frac{1}{p}(\omega_3-\omega_4)\right] \tag{66}$$

Thus in general

$$\omega_n=\frac{1}{\rho J_n}\left[T_n+K_{n-1}\frac{1}{p}(\omega_{n-1}-\omega_n)-K_n\frac{1}{p}(\omega_n-\omega_{n+1})\right] \tag{67}$$

Let one second real time $=\gamma$ seconds computer time. Then

$$T=\gamma t \tag{68}$$

and

$$t=\frac{T}{\gamma} \tag{69}$$

where T = computer time
and t = real time
Thus:

$$p(\ )=\frac{d(\ )}{dt}=\frac{d(\ )}{d\left(\frac{T}{\gamma}\right)}=\gamma\frac{d(\ )}{dT}=\gamma\rho(\ ) \tag{70}$$

Writing equation 67 in terms of computer time

$$\omega_n=\frac{1}{\gamma\rho J_n}\left[T_n+K_{n-1}\frac{1}{\gamma\rho}(\omega_{n-1}-\omega_n)-K_n\frac{1}{\gamma\rho}(\omega_n-\omega_{n+1})\right] \tag{71}$$

The application of Newton's second law to the torsional vibration damper shown in Fig. 8 gives (in terms of computer time).

$$K_1\frac{1}{\gamma\rho}(\omega_1-\omega_c)-K_2\frac{1}{\gamma\rho}(\omega_c-\omega_3)-\alpha(\omega_c-\omega_D)=J_c\gamma\rho\omega_c \tag{72}$$

$$\alpha(\omega_c-\omega_D)=J_D\gamma\rho\omega_D \tag{73}$$

If the moment of inertia of the can is neglected, equation 72 becomes:

$$K_1\frac{1}{\gamma\rho}(\omega_1-\omega_c)-K_2\frac{1}{\gamma\rho}(\omega_c-\omega_3)-\alpha(\omega_c-\omega_D)=0 \tag{74}$$

Rearranging terms

$$\omega_c=\frac{1}{\alpha}\left[K_1\frac{1}{\gamma\rho}(\omega_1-\omega_c)-K_2\frac{1}{\gamma\rho}(\omega_c-\omega_3)\right]+\omega_D \tag{75}$$

Rearranging equation 73

$$\omega_D=\frac{\alpha}{J_D\gamma\rho}(\omega_c-\omega_D)=\frac{\omega_c}{1+\frac{J_D\gamma\rho}{\alpha}} \tag{76}$$

Substituting equation 76 into 75

$$\omega_c=\frac{1}{\alpha}\left[K_1\frac{1}{\gamma\rho}(\omega_1-\omega_c)-K_2\frac{1}{\gamma\rho}(\omega_c-\omega_3)\right]+\frac{\omega_c}{1+\frac{J_D\gamma\rho}{\alpha}}$$

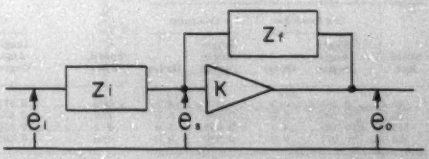


Fig. 16. Operational amplifier circuit

or

$$\omega_c = \left[1 - \frac{1}{1+\frac{J_D \gamma \rho}{\alpha}}\right] = \frac{1}{\alpha}\left[K_1 \frac{1}{\gamma \rho}(\omega_1 - \omega_c) - K_2 \frac{1}{\gamma \rho}(\omega_c - \omega_2)\right] \quad (77)$$

Basic Computer Equation[7]

Fig. 16 shows the typical operational amplifier circuit used in analog computer circuits. Z_i and Z_f are the input and feedback impedances respectively. These impedances are usually made up of resistors, capacitors, or combinations thereof. K (the triangle) is an extremely high gain d-c amplifier (a gain of 10^8 is not uncommon). The input impedance of the amplifier itself is very high, and consequently the grid current drawn can be assumed negligible (of the order of 10^{-10} amperes). The output impedance of the amplifier itself is of the order of a fraction of an ohm, and it also can be neglected.

Using the preceding assumption, and applying Kirchoff's law to the amplifier circuit shown by Fig. 16, the current through the input impedance must equal the current through the feedback impedance or

$$\frac{e_i - e_3}{Z_i} = \frac{e_3 - e_0}{Z_f} \quad (78)$$

The equation for the gain of the amplifier is

$$e_0 = K e_3 \quad (79)$$

where K = the amplifier gain

Combining equations 78 and 79, and rearranging:

$$\frac{e_0}{e_i} = -\frac{Z_f}{Z_i}\left[\frac{1}{\frac{1}{K}\left(\frac{Z_f}{Z_i}+1\right)+1}\right] \quad (80)$$

But since K is very large,

$\frac{1}{K}\left(1+\frac{Z_f}{Z_i}\right)$ is practically zero, and equation 80 becomes:

$$\frac{e_0}{e_i} = -\frac{Z_f}{Z_i} \quad (81)$$

References

1. Mechanical Vibrations (book). J. P. Den Hartog. McGraw-Hill Book Company, Inc., New York, N. Y., fourth edition, 1956, pp. 170–224.

2. *Ibid.*, pp. 23–78.

3. Practical Solution of Torsional Vibration Problems (book). W. Ker Wilson. John Wiley & Sons, Inc., New York, N. Y.; third edition, 1956, p. 350.

4. Harmonic Coefficients of Engine Torque Curves, F. P. Porter. *Transactions*, American Society of Mechanical Engineers, New York, N. Y. 1943, p. A–33.

5. Manual, Type IV Vibration Recorder. Industrial Manufacturing and Tool Company, Hazel Park, Mich.

6. Torsional Vibrations with Angular Backlash, W. A. Tuplin. *The Engineer*, London, England, vol. 160, 1935, p. 199.

7. Analog Computer Techniques (book), C. L. Johnson. McGraw-Hill Book Company, Inc., 1956, p. 9.

Ventilating Systems for Railway Traction Equipment

B. A. WIDELL
MEMBER AIEE

IN THE EARLY STAGES of the development of railway traction equipment the need for some form of ventilation became evident. This need increased with the continuing demand for increased output. As a result, cooling systems evolved from self- to forced ventilation. Their development was not, however, an unmixed benefit. Any ventilation tends to draw dirt, snow, moisture, and other contaminants into the machine, with harmful results. The undesirable effects may be reduced by design measures aimed at the following: 1. making the components more resistant to damage, 2. preventing the accumulation of foreign matter in critical locations, and 3. cleaning the air before it enters the machines. The proper measures to employ in a given instance are determined by design, practical and economic consideration.

Paper 58-507, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-ASME Railroad Conference, Cleveland, Ohio, April 9-10, 1958. Manuscript submitted December 26, 1957; made available for printing January 28, 1958.

B. A. Widell is with the General Electric Company, Erie, Pa.

Evolution of Ventilating Systems

Self-Ventilation

The early transit systems were operated most from a 600-volt trolley. As the transition from totally enclosed to fully ventilated motors took place, troubles occurred because of insulation grounds and rapid wear of the sides of brushes and carbon boxes. It was recognized early that moisture and dirt entering the motor were causing the grounds. To avoid the vacuum-cleaning action of sweeping up the streets that had resulted from some of the original schemes, the air intake covers were redesigned so that air was drawn from above the motor instead of at street level. Insulated baffles were made a part of the cover to provide a shield to deflect the incoming air away from the commutator and brush-holder; see Fig. 1.

Mummified field coils were developed during this period, and the amount of moisture-absorbing insulating material in all windings was reduced, thus adding materially to the life of the equipment under adverse operating conditions. Rapid wear of the carbon ways continued,

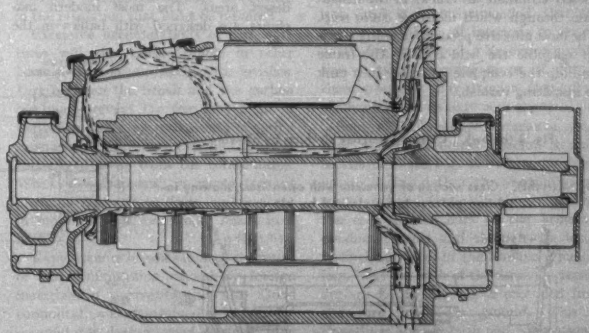


Fig. 1. Self-ventilated motor fitted with baffles to deflect incoming air away from the commutator and brush holders

Fig. 2. Accumulation of oily dirt completely blocking ventilating passages in motor armature

however, providing impetus for the development of the renewable carbon box.

The development of the PCC (President's Conference Committee) car for city transit in the 1930's marked another milestone in the march toward better motor protection against the elements. This car had four high-speed self-ventilated traction motors. They were connected to a common compartment by flexible bellows between the motors and air ducts. Air was taken into the compartment above the sills, thus providing the first modern clean-air ventilating system for traction motors on city cars. Later models of this car were provided with blowers so that fans on the motor shafts were no longer necessary. Under some conditions, however, the motors might operate through water. This could enter the air outlet holes in the motor frame. Therefore it was necessary to thoroughly seal the terminal connections of the field coils in order to completely eliminate all cracks in the insulation through which moisture could seep. The most effective process for doing this is to assemble the field coils in the frame and dip the complete assembly in a tank of insulating varnish.

Motors for application without a duct system are designed with insulation thicknesses, creepage distances, and varnish treatments that enable them to run creditably under most adverse conditions. The same features, however, generally cannot be omitted, even when the motor is supplied with clean air from a duct system. The additional cost of obtaining relatively clean air by installing air ducts to the traction motors will generally be more than offset by the lower motor maintenance costs.

Whether or not to use blowers or an air duct system for smaller, lower voltage machines is usually decided by the amount of maintenance involved in running the machine with self-ventilation and baffled covers. If the air holes through the armature laminations become plugged with dirt and must be cleaned out frequently, the expense of doing this may to some extent offset the cost of installing an air duct system. Traction motors which generate voltages higher than system voltages when used for dynamic braking are less susceptible to flashovers if provided with a clean-air duct system.

The experience gained with 600-volt motors was applied to the modern low-voltage axle-driven generators used for railway passenger car air conditioning. These were self-ventilated with good results except that dirt peculiar to some territories frequently caused the brushes to stick in the holders and the generator voltage would not build up. This difficulty was overcome by cutting slots in the carbon box adjacent to the brush and allowing the dirt to pass through. These machines, and early self-ventilated diesel-electric traction motors had to be protected against sand blasting which removes the insulation from the field coils when the equipment operates in sandy desert areas. The most modern machines are designed with baffles in the fan outlet chamber so located that the sand is directed away from the field coils; thus eliminating trouble from this source.

Use of Settling Chambers

Where space permits, much of the foreign material can be removed from the ventilating air by passing it through a settling chamber. The effectiveness of this system has been proved on multiple-unit cars operating from a 3,000-volt d-c overhead system. These cars are usually provided with self-ventilated motors having fans on the armature shafts. Air is drawn in at car roof level, passes through a settling chamber under the car body, and then into the traction motors. Much of the foreign material in the air is deposited in the settling chamber and clean ventilating air is delivered to the motors. Since this system does not remove moisture from the air, it is still necessary, in order to prevent grounds from this cause, to insulate the motor windings to withstand occasional periods of wetness. Another measure of the effectiveness of any air-cleaning system is the frequency of motor flashovers. On well-maintained 3,000-volt equipment, ventilated as described here, this is no more than is normal for the lower voltage systems.

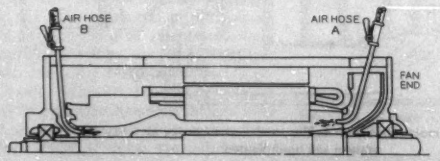
Forced Ventilation

With the trend in traction equipment toward higher horsepower ratings, more air is required for ventilation. The external blower with a duct system becomes increasingly desirable both for the cooling obtained and the possibility of taking the air far enough from the roadbed to assure its being relatively clean.

Separate blowers are used for ventilating the traction motors on all high-capacity electric and diesel-electric locomotives. The greater air flow thus obtained through the large low-speed motors

Fig. 3 (below). Cleaning armature air ducts by blowing

Fig. 4 (right). Cross section of generator with open risers, showing insulation surfaces to be cleaned by blowing



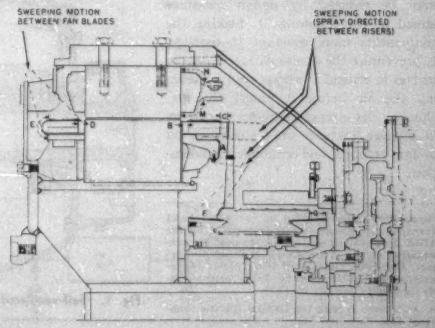


Fig. 5 (left). One method of cleaning open-riser armatures by means of an abrasive blast

Fig. 6 (right). Damage to field coil insulation resulting from blasting action of sand-laden ventilating air

permits higher ratings than if the machines were self-ventilated.

Today many diesel-electric locomotive generators are self-ventilated. Unfortunately they are so situated as to be subject to the smoke- and oil-laden air found in most locomotive engine rooms. Effective methods have been developed for periodic cleaning of the vital creepage surfaces of these generators. Improvements in insulation designs have frequently been found necessary to meet some of the worst operating conditions encountered. The most helpful solution would be to arrange the ventilating system so that oil-laden air could not reach the generator parts.

Problems Introduced by Ventilation

Ventilation is not an unmixed blessing. Cooling air entering motors and generators is never perfectly clean. More or less moisture, dirt, and other foreign material are entrained in the air stream and deposited in the equipment. Unfortunately, as the volume of air is increased more dirt is carried into the machines. The amount of contamination depends upon a number of factors, such as the type of ventilating system, amount of filtration, and the environment in which the locomotive is operating.

Effects of Dirt

Much has been written on the various maintenance methods used for cleaning traction motors and generators, and on the necessity for removing the dirty accumulations from the exposed insulations and other susceptible parts. If motors are not cleaned periodically, the mixtures of oil, water, and dirt will collect until an insulation check indicates that the level is no longer high enough to continue operation. By that time it is usually necessary to dismantle the machine and repair the part in trouble.

The insulation level of a machine is usually checked with a resistance-measuring device that subjects the insulation to a test up to 500 or 1,000 volts. The readings can be interpreted to decide whether a machine is fit for operation or must be worked on to bring up its insulation resistance.

When the windings of new machines are given insulation tests at the factory they are subjected to much higher voltage between turns and to ground than will be obtained in service. With use, since the machines are being subjected to oil, dirt, carbon dust, and water, it is to be expected that the insulation will be affected adversely. Then, when wet or extremely dirty, it would be possible to break down the insulation at a relatively low voltage. When low insulation resistance is found on a dry machine, the fault probably is caused by accumulations of dirt bridging the creepage surfaces of brushholders, armatures, and field coils. Proper cleaning of these surfaces should restore the insulation resistance to satisfactory levels. If it does not, there is probably mechanical damage somewhere in the insulation that must be repaired.

Excessive accumulations of dirt mixtures will also cause trouble in other parts of the machine, such as the brushes. Dirt between the brush and the carbon box may prevent the free movement of the brush and even overcome the spring resistance of the levers, causing the brush to be held away from the commutator. When this happens in generators it prevents the voltage from building up; in motors it will eventually result in a flashover.

Rapid wear of brushes is often the result of bad commutation caused by foreign materials in the ventilating air. Good commutation is maintained as long as the commutator has a good surface film. The presence of oil and dirt in the ventilating air may cause this film to break down, or the commutator surface to become glazed. Then poor commutation results.

Flashovers occasioned by atmospheric or environmental conditions are generally brought about by deterioration of the commutator film which is built up during favorable conditions. The breakdown of the film may be caused by small dirt particles imbedded in the commutator copper or in the brushes. These cause the brushes to vibrate and result in a flashover under adverse operating conditions. If the flash is severe enough the commutator must be smoothed before the machine can be put back into service.

Those machines which are ventilated by passing air through openings under the commutator and through longitudinal holes in armature punchings will receive a deposit of oil mixed with dirt in these passages which, if allowed to accumulate, would entirely shut off ventilation through this path; see Fig. 2. The armature windings will run hotter under this condition, and it is recommended that the passages be blown at intervals to get rid of the dirt; see Fig. 3. Excessive quantities or sticky accumulations of dirt in these passages may require the use of mechanical means to remove them.

Some of the creepage surfaces on open-riser machines require special cleaning techniques; see Fig. 4. These machines are so designed that the ventilating air passes by the vertical risers connecting the commutator bars to the armature conductor leads and through the longitudinal air ducts in the armature punchings. At the back of the risers a creepage surface (*A* in Fig. 4) is provided from armature leads to ground. This is directly in the air path and consequently becomes covered with the dirt carried by the ventilating air. Periodic removal of

this dirt film is necessary for successful operation, and one method of cleaning is illustrated in Fig. 5. Other creepage surfaces that also become covered with a dirt film are shown at points *B* through *N* on the longitudinal section in Fig. 4. These surfaces are designed to be readily accessible for cleaning.

Sand blasting of field coils is the result of local conditions, occurring mostly in desert country, where the ventilation system permits sand to enter with the air. Fig. 6 shows the exposed surface of bare copper conductor at the fan end of a generator field where the sand-laden air removed part of the insulation.

Protection Against Dirt

The various means adopted for protecting railway machines against contamination from foreign material in ventilating air may be grouped under three headings. The first is to seal the live parts of the machine so that they will be impervious to the contaminants in the air. The second is to baffle the air passages in the machine so only relatively clean air gets to the parts to be cooled. The third is to remove the contaminants from the air so that only clean air enters the machine. Some of the installations in the last class are of particular interest as they indicate the trend toward better cleaning systems.

Air Cleaning

Ventilating blowers are usually mounted inside the locomotive cab, and outside air is drawn through openings in the side sheets. In some designs a measure of screening is obtained by using louvers in these openings. A majority of diesel-electric locomotives employ carbody filters of the impingement type to clean the air entering the cab. Other designs of filters are also used for the same purpose.

Blowers for ventilating electric equipment on diesel-electric locomotives frequently draw air from the engine room. The purity of this air depends upon a number of factors, such as the type and condition of the filters, condition of the diesel engine, etc. Many modern blowers of both the radial- and axial-flow type are provided with air cleaners which are very effective in removing water and snow through the passages provided to the outside or bottom of the locomotive.

The centrifugal type of air cleaner is most adaptable to motor-generator sets; the air cleaner being mounted on one end and the fan drawing air through the set to the other end. This method of ventilation has been used extensively on motor-generator sets for 3,000-, 1,500-, and 600-volt systems. Some initial troubles were encountered on account of failure to provide enough protection from snow and moisture when the sets were shut down. Better enclosures were provided at the fan end of the machine and improved brush-holder insulators installed to ensure that the sets would not be handicapped even under the worst conditions.

Summary

The advantages of clean ventilating air for electric-traction equipment are obvious as far as increased reliability and reduced maintenance are concerned. Existing air-cleaning systems provide various degrees of protection against water and dirt in the cooling air, but still leave something to be desired. The choice of a system for a particular application should be based on economic and practical consideration.

Further improvement in the ventilating of traction motors and generators can be made in two directions: supplying cleaner air, and increasing the volume of air. Progress in these areas will be reflected in economic gains.

Field Experience with Electrostatic Precipitator on Diesel Locomotive Engine Air Intake

PELL KANGAS
NONMEMBER AIEE

W. C. KERN
NONMEMBER AIEE

E. L. RICHARDSON
NONMEMBER AIEE

Synopsis: An experimental electrostatic precipitator was specially designed to fit the limited space in a diesel locomotive. Its effectiveness as an engine intake air cleaner and its practicality as a regular component in the locomotive were evaluated after two years of heavy freight service.

IN DISCUSSIONS between representatives of the Westinghouse Electric Corporation and the Baltimore & Ohio Railroad Company the testing of the electrostatic precipitator was considered necessary as a part of the railroad's never-ending quest for eliminating from the air intake the abrasive dusts that ultimately cause considerable wear in the engine power assembly parts. It was hoped that the electrostatic precipitator with its high efficiency air cleaning ability would help to provide answers to the following questions, heretofore unanswered.

1. What is the extent of wear when all, or nearly all, road-dust abrasives are kept out of the engine and the engine is then subjected only to so-called mechanical wear?

2. How much longer will the engine operate between overhaul periods when the element of air-borne abrasives is eliminated?

The problem of designing an electrostatic precipitator that would meet the engine air intake requirements and fit the space available to it were solved the end of 1954. The installation was made in February 1955 and testing began in an area where the equipment was subjected to rigorous road service and was available nearly daily for required testing and inspection.

Concurrent with the testing of the electrostatic precipitator the railroad was also testing commercially available air filters of the oil-bath, air-cyclone, and the present standard impingement type which provided direct comparison with the electrostatic-type air cleaner.

During the first 6 months in field service the filter efficiency was studied and maintenance schedules were developed. It completed its 2 years of regular road service on the main line tracks of the Eastern Region.

Paper 58-505, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-ASME Railroad Conference, Cleveland, Ohio, April 9–10, 1958. Manuscript submitted November 29, 1957; made available for printing January 27, 1958.

Pell Kangas and W. C. Kern are with The Baltimore & Ohio Railroad Company, Baltimore, Md., and E. L. Richardson is with the Westinghouse Electric Corporation, Boston, Mass.

The power assemblies, consisting of cylinder liners, pistons, and piston rings, were measured at the completion of one year's service and the balance of the assemblies were measured at the end of 2 years.

The full potentialities of the electrostatic precipitator were not derived because of railroad operating and maintenance conditions. Some maintenance points, unfamiliar with the test equipment, were unable to make necessary repairs; consequently the precipitator was inoperative for short periods of locomotive operation. On several occasions the test locomotive was unable to return on time to its assigned maintenance terminal for regular 15-day maintenance.

The test was conducted on one *EMD* 16-cylinder 2-stroke cycle diesel engine having a bore and stroke of $8^1/_2$ by 10 inches, rated at 1,500 horsepower at 800 rpm.

Description of Electrostatic Precipitation

Two basic functions are usually performed by an electrostatic precipitator designed for high efficiency of dirt collection. The first function is the charging of the air-borne dirt particles; the second an electrostatic precipitation of the particles.

Dirt particles are charged by passing them through a field of unipolar ions. Two significant types of charging are usually considered to exist. They are designated as "field charging" and "diffusion charging" and act simultaneously but with degrees of effectiveness varying with particle characteristics. In general, it can be said that the quantity of charge a particle receives varies with the electric field strength, the concentration in the ion field and the time the particle is exposed to the field.

Charging of the particles requires the generation of ions. This is generally accomplished by two electrodes. One, the passive electrode, is a surface of relatively large area such as a plate or tube. The other is the active or emitting electrode of relatively small area, such as a fine wire. These electrodes are located so that when a potential difference is established between them the voltage gradient at the active electrode is high enough to cause ionization of the gas molecules near it. Ions of both signs are produced. Those of sign opposite to that of the active electrode travel to it. Those of like sign travel to the passive electrode across the intervening space and are available for particle charging. Precipitation is accomplished by interaction between the charge on the dirt particle and an electrostatic field normal to the direction of air flow.

There are two general types of this class of precipitator. The first type is the single stage in which both charging and precipitation are performed by a single electric field. An example of this type of precipitator would be a pipe with a small wire running axially through the center. Here the active electrode is at high voltage while the tube may be at ground potential. Ionization occurs around the wire and a radial flow of unipolar ions is established throughout the entire internal volume of the tube. Dirt particles passing through the tube are charged by the ions and the reaction between this charge and the electrostatic field established between the wire and tube cause the particles to precipitate on the inner walls of the tube.

The second type is the 2-stage precipitator. Here ionization and precipitation are accomplished in two stages. The ionization section, or ionizer, and the precipitator section, or collector, are separated. The ionizer may consist of a number of tubes or flat plates serving as passive or ground electrodes, spaced equally across the path of air flow, with small diameter wires strung between each pair. This stage functions mainly as an ionizer, although some precipitation occurs.

The collector stage consists of a number of flat plates equally spaced across the the air flow path, and orientated so that the dirt laden air will pass between adjacent plates of opposite potential, thus setting up a precipitating field. No ionization occurs. Oil is applied to the collector plates to retain the dirt which loses it charge when precipitated.

Design for Diesel Locomotive Application

Design Problem

The basic requirement was for an electrostatic precipitator to clean 5,400 cubic feet per minute of diesel engine intake air at high efficiency.

It was quickly realized, on inspection of the locomotive, that locating suitable space for the precipitator was problematic. The space must be large enough to accept the precipitator and allow reasonable facility for removal of its components for cleaning. In addition, the installation should not interfere with the inspection and maintenance of other equipment.

The largest available space was some distance from the engine air intake, but was finally ruled out because of difficulties that would be involved in installing air ducts.

Location of the unit over the engine was desirable because of the short connecting duct that would be required. This appeared impossible at first because of the very limited space. However, careful study of the situation indicated that best utilization of available space could be realized by the use of two precipitators, one on each side of the engine, connected to the engine blowers by a manifold. Therefore it was necessary to reduce the size of the units to handle air at more than twice the normal velocity.

Design Solution

Layouts were made to determine optimum dimensions of the apparatus. The final arrangement of a 2-stage precipitator is shown in Fig. 1. A manifold was attached to the two engine intake air blowers. A precipitator was attached to each end of a common plenum chamber and the whole assembly supported from the main roof hatch. This is shown in Fig. 2. A flexible duct connected the manifold to the plenum chamber.

On the final design it was necessary to consider other factors. The components to be removed for cleaning should not weigh over 50 pounds to allow handling by one man. Safety devices had to be provided for high-voltage protection of personnel. The unit must be sufficiently rugged to withstand vibration normally occurring in the engine.

Calculations were made to determine the basic design parameters. To attain the desired performance the ionizer was designed with a 20% increase in field strength and a 104% increase in ion concentration. Flat plate ground electrodes were used to reduce air velocity through the ionizer to a minimum. The collector plates were made longer in direction of air flow and spaced closer together to increase field strength 35%.

A model was built and tested in the laboratory at rated air velocity. Three types of aerosols were used to determine the collecting efficiency. It was found that with a monodispersed liquid aerosol of 0.3-micron diameter the efficiency was 95%. With normal atmospheric dirt, the efficiency was 96%. With Arizona Road Dust "fine" the efficiency was 98%. The Arizona Road Dust was used in the laboratory test as its particle size distribution was found to agree reasonably well with that of dust encountered in railroad service.

Efficiency was determined by a photoelectric device which indicated the relative dirt concentration in the air entering

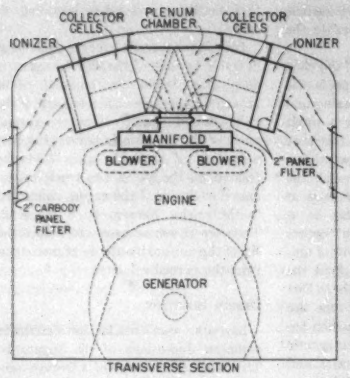


Fig. 1 (left). Arrangement of electrostatic precipitator on test engine

Fig. 3 (right). Collectors and ionizer removed from one end of the electrostatic precipitator

and leaving the precipitator. Efficiency on actual road test was determined by dust impingement on oiled microscope slides.

In the final design a casing was provided for the collector section. The collector, shown in Fig. 3, was made in two parts in order to reduce the component weight for easy handling. Each collector cell was fitted with quick acting fasteners to hold them in place. Electric connection to each cell was made by a spring type connector located at the rear.

The ionizer, shown in Fig. 3, was a single section arranged to clamp to the front face of the collector casing. Electric power supplied through an enclosed push-type connector. A wire screen over the face of the ionizer prevented accidental contact with high voltage parts. An interlock switch with time delay for capacitance discharge, required opening before access could be made to the unit and provided further personnel protection.

Fig. 4, the engine room, shows how the space over the engine air blowers was completely utilized by the electrostatic precipitator. As can be seen from Figs. 1 and 2, a 2-inch panel filter was attached to the front face of the ionizer and a similar panel filter was installed in the plenum behind the unit.

Electric power was supplied to the unit by a standard 100 volt 60 cycle commercial high-voltage power pack designed for 104 degree ambient and supply 12,800 volts direct current to the ionizer and 5,800 volts direct current to the collector cells. The power pack was fitted with resilient mounts for vibration protection. The power pack and a 110-volt 60-cycle 500-watt motor generator set operating from the locomotive's 64-volt direct current were mounted in the dynamic brake hatch.

The ionizer and collector cell were constructed almost entirely of aluminum. Flat plate ground electrodes were used. Thick electrodes would have reduced the effective area of ionizer and increased the air velocity through it. This would have resulted in lower particle exposure time in the ionizing field, and a decrease in efficiency.

The collector plates were supported on rods with spools to space them uniformly apart. The plate assemblies were clamped tightly together to insure rigidity and good electric contact between plates of like polarity. Slotted bars instead of spools could have been used for spacing the plates. Although this construction is lower in cost, the joints between the plates and bars are not tight, mechanical support is not good, the plates tend to be loose, and electric contact is indifferent. The main disadvantage is the burning of the plates at their support point. A collector

Fig. 2. Electrostatic precipitator supported from the main roof hatch of the diesel engine room

Fig. 4. View of electrostatic precipitator after installation in diesel engine room

cell may be considered to be a capacitor and a source of electric energy. High transient currents can exist within the cell when arc-over occurs. Burning of the plates can be very severe at joints of poor contact.

Test Method

Air Cleaning Efficiency

The method for measuring the air cleaning efficiency of the electrostatic precipitator has been previously described.[1] This method was specially developed for this test and utilized oil-coated 1 by 3 inch microscope slides placed in the air stream, both in front of and behind the electrostatic precipitator. Air-borne particles were impinged on the slides and particle size distribution and particle counts were made with a microscope at 400X magnification. This provided means for examining particles ranging in size down to one micron.

This method fulfilled the need for performing the efficiency tests on the electrostatic filter in regular service and thereby studying the affect of time on efficiency. The microscope slides were permitted to ride from 80 to 300 miles before their removal for study of the particle size distribution. The efficiency of the electrostatic precipitator was therefore determined as the ratio of the number of particles impinged on the front and rear slides per unit area or average field on the slide.

Wear Determination

The method for measuring wear was based on having all 16 power assemblies measured according to the engine manufacturer's recommended procedure prior to installation. The parts measured were:

1. Cylinder liner diameters both parallel to and perpendicular to the crankshaft at $1^9/_{16}$ inches and 6 inches from the top.
2. Piston ring gap.
3. Piston ring groove width.
4. Piston ring thickness.

Cylinder diameters were determined before and after test by the use of a dial bore gage accurate to 0.0001 inch. Piston ring gap was measured by placing each ring in a master ring gage also accurate to 0.0001 inch and measuring the gap with a taper gage graduated in thousandths of an inch. Piston ring grooves were measured for width by a series of plug gages. Ring thickness was measured with 1 inch outside micrometers. All measuring work was performed by test personnel.

After one year of service, power assemblies no. 3 and no. 11 were removed for measurement in order to get an indication of the wear rate at this time. The remaining 14 assemblies were measured at two years.

Field Test Experience

Type of Service

The diesel locomotive with its test filter equipment was placed in heavy freight service over a 135-mile section of hilly branch-line railroad. It was held captive in this territory for 6 months making about 175 miles per day. During the remaining 18 months of testing it operated in mainline freight service, principally over the mountainous section of the eastern region under extremely dusty atmospheric conditions.

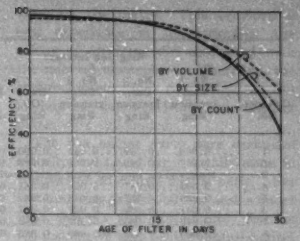


Fig. 5. Efficiency of electrostatic precipitator

Routine Maintenance

Maintenance of the electrostatic precipitator was relatively simple since two complete sets of collector cells and ionizers were used. One set being in use while the dirty set was held in the shop for cleaning and treatment. The time required for two men to change out the electrostatic precipitator parts was about 20 minutes. The parts were not too large for one man to handle; however, it was safer when two men performed this work.

Cleaning of the dirty ionizer and collector cells was done either by use of a steam jet gun drawing soap solution designated for cleaning aluminum parts or by dipping in wash tank and then rinsing clean with water. When dry, the collector cells were dipped in a filter treating compound in order to assure developing a mechanical bond between the dirt particles and the collector cell plates.

Maintenance Problems

No trouble was reported during the test period due to blow-off dirt caused by vibration or electric arc-over between the

Table I. Test Engine Wear Data

Cylinder No.	Ring Gap Increase					Ring Side Wear			Ring Groove Wear			Cylinder Liner Wear			
	No. 1 Compression Ring	No. 2 Compression Ring	No. 3 Compression Ring	No. 1 Oil Ring	No. 2 Oil Ring	No. 1 Compression Ring	No. 2 Compression Ring	No. 3 Compression Ring	No. 1 Compression Groove	No. 2 Compression Groove	No. 3 Compression Groove	A	A1	B	B1
1	Destroyed by mistake before measurement								0.0030	0.0010	0.0010	0.0032	0.0039	0.0035	0.0013
2	Destroyed by mistake before measurement								0.0020	0.0010	0.0010	0.0050	0.0038	0.0047	0.0032
3	Removed for one-year wear determination														
4	Destroyed by mistake before measurement								0.0030	0.0010	0.0005	0.0055	0.0040	0.0036	0.0026
5	0.035	0.018	0.017	0.021	0.020	0.0010	0.0000	0.0000	0.0030	0.0015	0.0010	0.0073	0.0031	0.0058	0.0014
6	0.027	0.018	0.014	0.026	0.023	0.0005	0.0000	0.0000	0.0020	0.0010	0.0000	0.0043	0.0038	0.0032	0.0008
7	Broke	0.019	0.014	0.023	0.023	0.0005	0.0000	0.0000	0.0020	0.0010	0.0010	0.0056	0.0043	0.0046	0.0021
8	0.030	0.018	0.016	0.030	0.023	0.0000	0.0000	0.0000	0.0025	0.0010	0.0015	0.0045	0.0049	0.0041	0.0016
9	0.034	0.021	0.017	0.022	0.020	0.0000	0.0000	0.0000	0.0030	0.0020	0.0015	0.0046	0.0039	0.0043	0.0038
10	0.043	0.027	0.022	0.030	0.026	0.0000	0.0000	0.0000	0.0040	0.0010	0.0010	0.0071	0.0052	0.0065	0.0012
11	Removed for one-year wear determination														
12	0.037	0.023	0.018	0.028	0.015	0.0000	0.0000	0.0000	0.0020	0.0010	0.0010	0.0057	0.0041	0.0057	0.0033
13	0.029	0.017	0.014	0.025	0.018	0.0000	0.0000	0.0000	0.0030	0.0010	0.0005	0.0032	0.0046	0.0037	0.0013
14	0.036	0.017	0.016	0.027	0.020	0.0000	0.0000	0.0000	0.0030	0.0015	0.0015	0.0034	0.0029	0.0062	0.0002
15	0.050	0.027	0.023	0.025	0.023	0.0000	0.0000	0.0000	0.0030	0.0000	0.0000	0.0084	0.0051	0.0068	0.0037
16	0.040	0.023	0.019	0.024	0.C22	0.0000	0.0000	0.0000	0.0030	0.0010	0.0005	0.0056	0.0042	0.0060	0.0018
Average	0.036	0.021	0.017	0.025	0.021	0.0002	0.0000	0.0000	0.0027	0.0011	0.0008	0.0047		0.0034	

A—Perpendicular to crankshaft at $1^9/_{16}$-inch level.
A1—Parallel to crankshaft at $1^9/_{16}$-inch level.
B—Perpendicular to crankshaft at 6-inch level.
B1—Parallel to crankshaft at 6-inch level.

Table II. Control Engine Wear Data

Cylinder No.	Ring Gap Increase					Ring Side Wear			Ring Groove Wear			Cylinder Liner Wear			
	No. 1 Compression Ring	No. 2 Compression Ring	No. 3 Compression Ring	No. 1 Oil Ring	No. 2 Oil Ring	No. 1 Compression Ring	No. 2 Compression Ring	No. 3 Compression Ring	No. 1 Compression Groove	No. 2 Compression Groove	No. 3 Compression Groove	A	A1	B	B1
1	0.066	0.049	0.041	0.090	0.078	0.0025	0.0008	0.0005	0.0045	0.0025	0.0015	0.0085	0.0051	0.0086	0.0039
2	0.107	0.065	0.069	0.088	0.082	0.0230	0.0020	0.0005	0.0050	0.0015	0.0005	0.0066	0.0080	0.0067	0.0021
3	Removed for one-year wear determination														
4	0.062	0.037	0.027	0.081	0.076	0.0045	0.0010	0.0005	0.0025	0.0010	0.0010	0.0081	0.0057	0.0067	0.0005
5	0.039	0.037	0.025	0.082	0.068	0.0030	0.0005	0.0005	0.0057	0.0010	0.0010	0.0037	0.0032	0.0020	0.0003
6	0.041	0.030	0.026	0.087	0.068	0.0035	0.0010	0.0005	0.0065	0.0030	0.0010	0.0103	0.0032	0.0075	0.0039
7	0.045	0.030	0.027	0.085	0.072	0.0060	0.0000	0.0005	0.0020	0.0010	0.0010	0.0054	0.0079	0.0054	0.0004
8	0.065	0.041	0.040	0.085	0.068	0.0030	0.0005	0.0000	0.0030	0.0010	0.0010	0.0059	0.0042	0.0059	0.0003
9	0.140	0.084	0.074	0.084	0.078	0.0150	0.0090	0.0010	0.0045	0.0010	0.0000	0.0133	0.0086	0.0118	0.0010
10	0.133	0.093	0.075	0.091	0.075	0.0080	0.0000	0.0005	0.0085	0.0030	0.0020	0.0168	0.0098	0.0127	0.0012
11	Removed for one-year wear determination														
12	0.043	0.029	0.026	0.083	0.083	0.0020	0.0000	0.0000	0.0060	0.0010	0.0010	0.0049	0.0042	0.0085	0.0020
13	0.049	0.028	0.029	0.082	0.065	0.0010	0.0005	0.0000	0.0040	0.0015	0.0015	0.0059	0.0044	0.0069	0.0008
14	0.085	0.047	0.041	0.062	0.063	0.0011	0.0000	0.0005	0.0040	0.0010	0.0010	0.0094	0.0072	0.0089	0.0019
15	0.067	0.039	0.034	0.071	0.077	0.0010	0.0000	0.0000	0.0030	0.0010	0.0010	0.0082	0.0061	0.0081	0.0006
16	Removed during test due to water seal leak.														
Average	0.072	0.047	0.041	0.082	0.073	0.0056	0.0012	0.0004	0.0047	0.0015	0.0010	0.0071		0.0045	

A—Perpendicular to crankshaft at 1 9/16-inch level
B—Perpendicular to crankshaft at 6-inch level
A1—Parallel to crankshaft at 1 9/16-inch level
B1—Parallel to crankshaft at 6-inch level

collector plates. It was noted that even rather heavy hammering of the cells with the fist failed to dislodge any dirt from the plates.

Very little trouble was encountered at the maintenance shop. Air leakage was noted at the joint between the 2-inch panel prefilter and the ionizer and was corrected. Although the electric connectors on the cells were fastened by a single bolt and had a tendency to rotate out of alignment, it was found that these bolts could be tightened sufficiently so that the trouble was eliminated. However, future designs should have better fastening means.

The insulators used on the collector cells were of a type ordinarily used in stationary applications. They are prevented from rotating in their seats by a small flat spring. Vibration caused several of these springs to fall out of place. Their loss did not affect the operation of the unit and they were replaced when found missing.

Two surge limiting resistors in the power pack failed and had to be replaced. Failure of these resistors was not believed to be due to vibration or heat. Compound was noted oozing from beneath the cover of the high voltage transformer, apparently due to high ambient temperature.

Two troubles were reported with ionizer insulators.

1. Insulators were broken apparently by dropping the ionizer or by malfunctioning of the cable connector.

2. Corrosion products around the insulator caused a short circuit. Insulators of a more suitable design were installed. Also during the period several ionizer wires broke and were replaced.

Bearing failure on the d-c motor of the motor generator supplying power to the precipitator power pack resulted in the precipitator being inoperative for 15 days. On one occasion the locomotive arrived in the shop with a report that the precipitator was not operating. Inspection revealed that the motor generator set was not running due to a raised brush on the d-c motor.

It was reported at the shop that the precipitator required frequent maintenance of a minor character. On several occasions the units failed to function on reassembly after cleaning and had to be taken down. The frequency of this difficulty decreased with experience. It was also reported that on ten occasions the locomotive returned to the shop with the precipitators not operating. In four of these cases failure was due to the motor generator set. The cause for the other cases of shutdown in unknown as on each instance the unit was checked and was found to operate properly. No apparent reason for the power source being shut

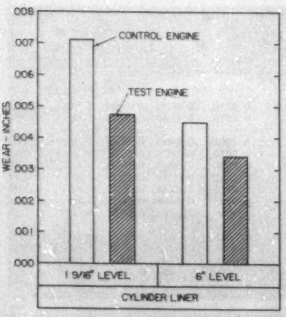


Fig. 6. Cylinder liner wear

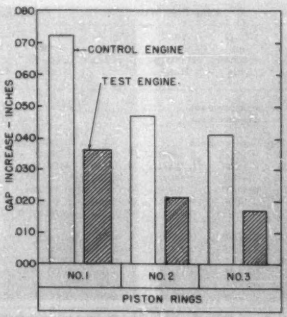


Fig. 7. Piston ring gap increase

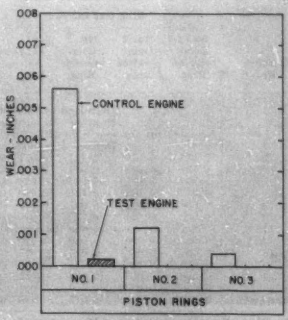


Fig. 8. Piston ring side wear

down could be found that would render the precipitator inoperative.

Test Results

Precipitator Efficiency

The average efficiency curves for the electrostatic precipitator are shown in Fig. 5, for a 30-day period. The curves are based on particle count, size and volume. The curves, by size and volume, are derived by computation from the particle size distribution data.

The tests indicate that with oil-treated collector plates the efficiency by count ranged from the initial efficiency of 98% to about 93% when the collectors had been in service 15 days and after 30 days the efficiency had decreased to about 40%. This drop in efficiency made it necessary to change out the collector cells and ionizers every 15 days in order to maintain a high efficiency. This changeout schedule conformed to shop practice of semimonthly inspections of engines.

Test runs made with untreated or dry collector plates did not show a significant difference in efficiency with those that were coated with a tacky oil. The deposits on the collector plates in both cases were similar in appearance and subsequent analysis showed them to have coatings similarly composed of air-borne dirt, carbonaceous material, and over 50% by weight of oil. It was evident that the electrostatic precipitator was collecting the oily vapors of the diesel engine room and the oily carbonaceous products that are generally present in railroad atmospheres.

Wear Measurement

The 2-year wear data for the test engine is compared with wear data derived from the control engine equipped with the standard 4-inch impingement-type panel filters. These data are given in Tables I and II. Two power assemblies were removed from the engine and measured after one year of service in order to get an indication of the wear rate for this period.

The average cylinder liner wear for the remaining assemblies (Fig. 6) shows a reduction of 34% at the $1^{9}/_{16}$-inch level and 25% at the 6-inch level for the test engine compared to the control engine. In terms of out-of-round wear, only two test engine cylinder bores were found to exceed the limit of 0.005 inch compared to 12 from the control engine.

The two out-of-roundness values, exceeding the tolerance limit of 0.005 inch, for the test engine at the 6-inch level of the cylinder liners were 0.0053 inch and 0.0060 inch, resulting in an average figure of 0.0056 inch. On the control engine the out-of-roundness ranged from 0.0046 inch to 0.0115 inch and the average of the 12 values was 0.0067 inch.

Compression ring wear, measured as gap increase, showed an average reduction of approximately 54% for the test engine. Oil ring wear was further reduced by the test engine by as much as 70%. This is shown in Fig. 7. Average oil ring gap increase for the control engine exceeded that for the no. 1 compression ring by 6%, whereas for the test engine oil ring wear was less than that for no. 1 compression ring wear by 37%. Ring side wear for the test engine was 96% less than that for the control engine as shown in Fig. 8.

Discussion

The nature of most maintenance problems encountered during the testing of the electrostatic precipitator, although minor in character, would undoubtedly be eliminated by improvement in materials and design. It was generally felt that the prime objective for any future design would be to extend the high degree of efficiency obtained during the initial 15-day period to a service period lasting 30 days.

A good indication of the air cleaning efficiency of the electrostatic precipitator was apparent when an inspection of the plenum chamber showed no dirt deposits. All interior surfaces of the plenum chamber had not been cleaned during the period of the 2-year test.

An interesting finding was developed in this test in that the dirt deposits on the collector plates, as well as on the ionizer plates, adhered tenaciously to the plate surfaces. Evidently the precipitator collected all air-borne particles including oily and carbonaceous components of the railroad atmosphere. Such deposits could not be removed by pounding or vibrating the plates. This provided a new experience in understanding the nature of air-borne dirt associated with the railroad atmosphere. Therefore, the earlier apprehensions concerning the unloading of deposits from the collector plates were unfounded.

The test engine parts were examined visually and found exceptionally clean in comparison to control engine parts. For instance, ring groove, oil ring and cylinder liner deposits were considerably less for the test engine than for the control engine. The rate of engine deposits, therefore, seemed to correlate with parts wear. It was generally agreed that the test engine could have continued to operate efficiently well beyond the normal 2-year overhaul period based on the wear rates experienced.

Conclusions

The first experimental design and application of an electrostatic precipitator as a diesel locomotive engine air cleaner in regular railroad service for 2 years was successful, in that,

1. Wear of engine parts was significantly less than that for the control engine.
2. Maintenance was not difficult and conformed to normal practices used in the railroad shop.
3. The electrostatic precipitator operated as installed without need for making any major changes.
4. Parts failures were minor in character and were corrected by shop personnel.

Field experience with the electrostatic precipitator indicated the following problems:

1. The progressive reduction in particle collecting efficiency during the first 15 days from 98% to 93%, by count, was largely related to high air velocities through the collector cells.
2. Collector efficiency dropped rapidly after 15 days of service.
3. Due to the limited air space in front of each precipitator unit the air velocities through some sections of the ionizer and collector were excessive.

Reference

1. Air Filtration and Diesel Engine Wear, Paul Kangas. *S.A.E. Journal*, New York, N. Y., June 1956, pp. 132–38.

Design of a Drive System for a Rotating Arm

H. E. PRUCHA
NONMEMBER AIEE

THE SELECTION and preliminary design of a drive system for a unique test facility called a "rotating arm" was carried on simultaneously with the structural and mechanical designs. This rotating arm is required for directional-stability and maneuverability studies of high-speed submarines and surface ships. Large-scale models will be towed in circular paths with radii ranging from 12½ feet to 120 feet and at submergence levels of as much as 10 feet. Final detail design of the electric machinery and control circuitry is contingent upon the results of exceedingly painstaking analog computer studies now partially completed. These studies are required of the manufacturer as proof of the adequacy of the machinery and its associated control circuitry for the intended application.

Fig. 1 is a partial tracing of an actual contract drawing and illustrates the essential features. The primary structure is an aluminum truss 129 feet long, 20 feet wide, with a maximum height of 20 feet. The inner end of the primary structure pivots on a large roller bearing supported by a semiconical island. Fig. 2 shows the island shortly after the concrete forms were removed. The skirt about the upper cylinder is a beach that helps prevent reflections of waves. The outer end is supported by wheels which travel on a circumferential track. Other tracks, attached to the bottom of the arm, support a wheeled boxlike structure to which model-supporting struts are fastened. A windlass and cable permit location of the model, together with its supports, at any desired radius and it is then securely clamped in place. The model support carriage contains a turntable so that the model can be oriented at any desired angle within 30 degrees of a tangent to the circular path. Other positioning equipment will allow tilting of the model-supporting struts if other than the vertical position is required.

The model with its struts may weigh as much as 16,000 pounds. A 2,600-pound model located 16 feet below the bottom chord of the arm will have a maximum drag of 1,700 pounds at the 50-foot radius when the arm is revolving at top speed. Requirements for high natural transverse vibration frequencies, in both the horizontal and vertical planes, and for a high natural torsional frequency, together with a stipulation that the maximum change in model attitude be no more than ±1 degree in any direction from start to finish of a series of tests, resulted in the selection of the massive trussed design of tubular members that comprises the primary arm. To keep inertia as low as possible, aluminum tubing was selected as the construction material. A wind-resistance analysis showed the advantage of providing fairings for the major structural members at the outer third of the structure.

Some of the data pertaining to the performance of the hydrodynamic models are derived much in the fashion of aerodynamic models in wind tunnels; that is, force- and moment-measuring dynamometers are part of the support system at the point of attachment to the models. The accurate determination of such data requires, among many other conditions, that model attitude, depth of submergence, and test speed all be maintained within very close limits. A variation in speed of a hydrodynamic model with respect to the water will cause greater errors in the measurement of drag forces than with an equivalent variation of air speed with respect to an aerodynamic model. The ratio of drag force to the mass of the hydrodynamic model is usually very small, therefore, the acceleration force required to change speed of the model, even a fraction of a knot, is very sizable. As an example, a typical 20-foot 2,480-pound submarine model, when towed at 6 knots, will have a drag force of about 47 pounds. If the speed of this model is increased by as little as 0.02 knot in 1/2 second, the force required to accelerate the model would normally be calculated as about 0.002 g or 4.96 pounds; however, an additional accelerating force is registered because of the phenomenon known as the virtual entrained mass of water. In the case of a submerged submarine model being moved forward, this effectively adds 10% to the actual mass of the model. Thus, it can be seen that slight changes in speed cause serious errors in drag measurement. Although it is possible to correct data for speed changes, it is less laborious, less costly, and more accurate to hold speed as constant as possible within economic feasibility. Established policy, therefore, is to design drive systems so as to take advantage of the highest accuracies obtainable in commercially available equipment. Past experience has shown that speed-regulated systems can be so designed as to have maximum speed errors ranging from ±1/10 to ±1/20% of the top speed specified. For this application, the final specifications allow a maximum error of ±0.000425 radian per second at any speed up to 0.425 radian per second and a maximum error of ±1/10% of any speed above 0.425 radian per second.

Steady-state arm speeds as high as 30 knots at the 120-foot radius must be attained within 180 degrees of movement from rest. This corresponds to a rotational speed of 0.425 radian per second. The high acceleration rate is required so as to permit certain types of tests in the remaining 180 degrees of undisturbed water. An additional requirement is that a maximum test speed of 0.71 radian per second be attainable within 2½ revolutions of the arm. Reverse rotation up to 0.212 radian per second is also required to facilitate the work but not as a precisely controlled test speed.

Although every known type of drive received some consideration as to feasibility, comparative cost, and other factors, it was apparent quite early that a Ward Leonard type of system would prove to be the best choice. However, experience with another Ward Leonard system on a totally different kind of test facility indicated that an allowance of about 3 seconds would have to be made for transients to die within acceptable limits after initial acceleration. Therefore, acceleration to 0.425 radian per second must actually take place in about 90 degrees of movement.

The first weight and size estimates and calculations indicated that a peak torque of at least 1.4 million pound-feet must be applied about the pivot point of the arm in order to accelerate it sufficiently within 90 degree of movement.

With the first attempts to develop this new facility, an investigation was made of all known existing or planned facilities which, however faintly, resembled the

Paper 58-559, recommended by the AIEE Marine Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Washington, D. C., April 28-30, 1958. Manuscript submitted January 16, 1958; made available for printing March 25, 1958.

H. E. PRUCHA is with the U. S. Navy Department, David Taylor Model Basin, Washington, D. C.

contemplated rotating arm. With the single exception of a much smaller non-regulated arm built in Paris, France, about 15 years ago, central drive systems are used. The largest of the group is a human centrifuge built at the Naval Air Development and Material Center at Johnsville, Pa. This centrifuge has a counterbalanced arm 50 feet in length. This arm supports a positionable gondola, the purpose of which is to study the behavior of subjects exposed to programmed acceleration forces ranging up to 35 g. It is required, therefore, that this arm be capable of rapid acceleration to a speed of 4 or more radians per second. Because arm deflections and natural frequencies do not materially affect test results, the arm was built of smaller tubular members and, in general, is very much smaller than the Model Basin Facility. A huge, vertically mounted 4,200-hp (horsepower) d-c motor serves both as the support and the drive means for the centrifuge, the radius arm being directly fastened to the armature shaft. This motor is so massive that a railroad siding had to be constructed at the site and a special high-lift traveling crane provided for its erection and servicing. Such handling facilities are expensive but when additional limitations, such as locating the drive in a 20-foot-diameter island in the middle of a 130-foot-diameter 20-foot-deep basin are added, this approach becomes rather impracticable. Although at first the functional specifications permitted a 20-foot-diameter island, this was later reduced to a 6-foot-diameter island extending several feet below the surface of the water and surmounting a cone-shaped support.

There are a few other rotating arms built for various purposes and though these are all very much smaller than the Model Basin Facility, they have similar drive arrangements. In general these arms are supported by cantilevered vertical shafts which are driven by horizontal motors acting through one or more stages of reduction gearing. If the contemplated drive had followed this pattern, a vertical shaft, 6 feet in diameter, would have been required, provided an adequate top bearing support could be designed. Then, to provide enough space to house the reduction gearing and the motor, it would be necessary to excavate a large chamber and access passage below the basin floor. Because of the cost of excavation, the danger of flooding the machinery, and the difficulty of removing major pieces of equipment for repairs, this approach to a drive system was abandoned.

Another variation which was seriously considered was a center drive scheme which would require a number of planetary gears, individually driven, and a very-large-diameter gear (11 feet or so) which would be fastened to the island. In this arrangement, the planetary gears would be supported in proper relationship to the large gear by a massive gear case. This gear case would also serve to support one end of the arm. A preliminary analysis showed that deflections in the gear teeth, twist in the island, wind-up in the motor shafts, and whipping of the outer end of the arm were likely to cause intolerable low-frequency oscillations (2 cycles per second or so). A variety of ways of adding damping to the outer end of the arm were briefly considered and found to be impracticable. Other objections to the planetary gear drive were: excessive cost of special gears and case, danger of oil leaking into basin water, and the necessity of emptying the basin and virtually dismantling the arm to effect major gear repairs. It became evident that the point of application of the drive torque would have to be at the outer end of the arm.

Other facility drives in which variable-displacement pumps, powered by synchronous motors, feed piston-type hydraulic motors, have proved to be highly satisfactory. However, in these previous applications there was no need for rapid acceleration to high speed, hence the torque ratio to total mass is low and maximum accelerations are limited to about 0.1 g as contrasted to the 0.4 g required for the Model Basin arm.

One arrangement of a hydraulic drive, briefly considered and found unsuitable, consisted of one or more synchronous motors and pumps located close to the center island with hydraulic motors driving traction wheels at the periphery of the arm. This system would have required oil supply and delivery lines, each at least 100 feet long. Analysis quickly demon-

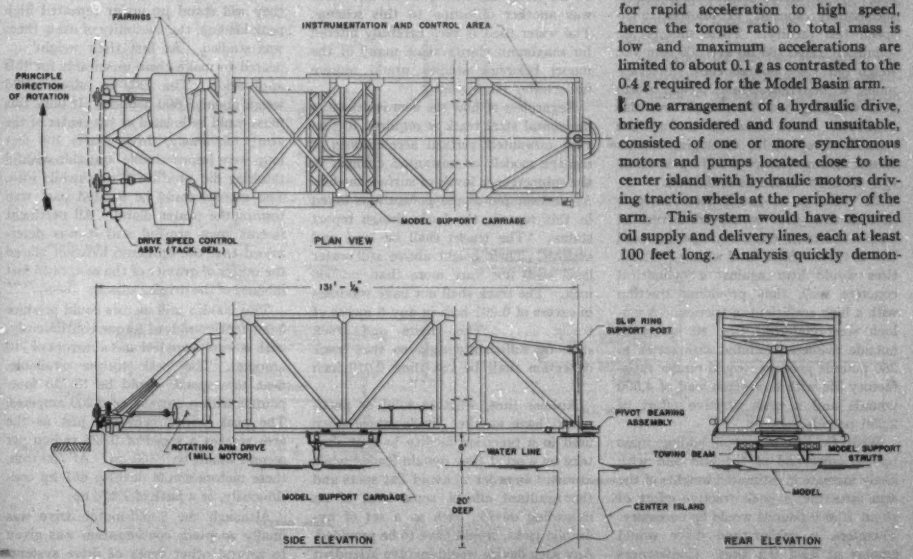


Fig. 1. Rotating arm assembly

Fig. 2. Rotating arm center support (shown in right foreground), note size compared to workman. Roof arches rise about 86 feet over floor of excavation

strated that the elasticity of such long lines, together with the compressibility of the oil, would quite likely cause intolerable oscillations of the arm. Other objections were to the waste of valuable space at the center of the arm to house the necessary equipment and the high cost of the hydraulic system.

A Ward Leonard system utilizing four d-c traction motors and worm reduction gears was investigated next. The particular type of motor considered had been successfully applied in another high-speed test facility. These motors are nominally rated 165 hp at 3,500 rpm and weigh but 1,300 pounds each. Furthermore, a matching Hindley-type worm gear weighing but 700 pounds was also available. These traction motors are capable of repeatedly developing a peak torque corresponding to 360 hp at 3,500 rpm without undue deterioration.

In the first proposal, these traction motors and gears would drive vertical output shafts on which would be mounted two or three rubber-tired wheels. These tires would bear against a cylindrical concrete wall, thus providing traction with a high coefficient of friction. A 9.5-inch steel-reinforced tire, 40 inches in outside diameter, inflated with water to 280 pounds pressure, would render satisfactory life with a nominal load of 4,500 pounds and a peak tractive effort of 2,250 pounds.

With the drive system adding mass to the outer end of the arm and with likely increase in estimated weights of the arm structure, a peak tractive effort of about 20,000 pounds would be necessary. Therefore, a rubber-tired drive would employ at least ten tires. Conferences with tire manufacturers offered no hope whatever of operating satisfactorily at higher tire loadings than those cited.

If four matched drive motors were used, each vertical shaft would have to carry three tires. This meant that the cylindrical track would have to be about 4 feet high. Preliminary estimates of the cost of an accurately formed concrete cylinder, about 130 feet in inside diameter, were discouraging. The possibility that the concrete would form a dust that would cloud the water in the test basin was another objection to this scheme. The water used is very carefully filtered for maximum clarity since many of the model behavior studies would require underwater photography.

Regardless of how the arm is driven, a horizontal steel track is required to prevent unwanted vertical accelerations of captive models or unwanted changes in the submergence levels of surface models. As an example of the precautions taken in this regard, the basic design report states: "The tracks shall be level and smooth. Their height above still water level shall not vary more than ± 0.006 inch. The track shall not have waviness in excess of 0.001 inch in any 6 inches of track. .....The chairs and track shall be sufficiently rigid so that track deflection shall be less than 0.010 inch ....."

Rubber tires, whether solid or pneumatic, have another disadvantage when used in a precision facility in that they take on a set if they remain loaded while at rest. In order to avoid flat spots and the resultant effects, some kind of an unloading device, such as a set of hydraulic jacks, would have to be provided. Any such device would involve attendant complications.

With the need for the steel guide wheels and tracks established, consideration was given to the feasibility of using these for the drive function as well. The effective radius at which these drive wheels would operate was now known to be approximately 130 feet. Previous experience with steel drive wheels indicated that a size ranging from 30 to 40 inches would probably be suitable.

Subsequent analysis showed that five traction motors, capable of developing a peak of 1,800 hp, would be needed. Since each armature would require 230 volts and since a series system was deemed essential to assure proper load sharing the system voltage would be 1,150 at a top speed, thus necessitating 1,500-volt switchgear and wiring. Peak current for these five motors would be about 1,600 amperes. These values of current and voltage were not too objectionable except for the difficulty of providing a mid-point ground.

At about this stage of progress the suggestion was made that lower speed motors might be suitable, especially if reduction gearing could be eliminated. A 30-inch drive wheel would rotate at about 420 rpm for an arm speed of 0.425 radian per second and at about 700 rpm for an arm speed of 0.71 radian per second. Since mill motors are especially designed for high continuous torque output and since they will stand up under repeated high peak loading, the feasibility of using them was studied. At first their weight appeared to make them unsuitable for this application. The 200-hp size would weigh about 8,000 pounds. If only this mass could be located at the center of the arm! Obviously, drive shafts 100 feet long were impracticable, but this wishful thinking did produce a worthwhile idea. The motors could be located part way toward the center island. All pertinent factors were studied and it was determined that the optimum location placed the center of gravity of the motors 15 feet inboard of the driving wheels.

Two 200-hp mill motors could produce 5,000 foot-pounds of torque continuously, with 460 volts applied and a current of 710 amperes. The peak torque available, near zero speed, would be 13,750 foot-pounds with a current, of 2,200 amperes. The peak power delivered just as the arm reaches a speed of 0.425 radian per second would be 1,100 hp. At 700 rpm, these motors would develop 667 hp continuously, or a peak of 1,375 hp.

Although the 2-mill-motor drive was finally adopted, consideration was given to several other types of drive systems that were known to perform well over at

least a 20-to-1 speed range. One such system would comprise induction motors fed with variable-frequency power. This power would be produced from standard 60-cycle power by means of a frequency converter. There are numerous ways of controlling the speed of a frequency converter. For frequencies below 60 cycles, mechanical energy is taken off the converter shaft and absorbed by an eddy-current brake or by a d-c machine. To increase the operating-speed range it is necessary to produce frequency above 60 cycles, which makes it necessary to put mechanical energy into the shaft of the converter. The simplest variable-frequency system that offers sufficient speed range would require a synchronous drive motor, an eddy-current slip coupling, and an eddy-current brake, together with excitation equipment for each of these units.

When cost estimates, which included sufficient substation volume, machinery foundation costs, etc., were compiled, the d-c mill motor system, powered Ward Leonard fashion, was next to the lowest in price. The only system that would probably cost less would be one using grid-controlled mercury-arc rectifiers. These have been used to power printing presses which operate over a wide range of speeds, but reliable operation with very precise speed control has not been demonstrated. It is believed that these rectifiers would prove satisfactory over a 4-to-1 speed range, and that this range could be extended by a factor of three or four by use of anode transformers with taps at several voltages.

Since some dynamic braking, reverse speed up to 0.212 radian per second, and the ability to spot the arm precisely for ease in attaching models, are prerequisites, the mercury-arc rectifier would have to be supplemented by an auxiliary rotating generator plus a complex switching scheme.

It was finally concluded that the straight Ward Leonard system, with tachometer feedback regulation, was the most suitable and least costly system available.

Smooth and accurately ground rails and wheels are essential for these facilities. Wheel slippage would cause scoring of the track surfaces, so every effort has been made to prevent this. Consequently, peak tractive effort must depend on the very low coefficient of friction of 0.15.

Each 200-hp mill motor developing a peak torque of 11,450 pound-feet on a 30-inch-diameter wheel will produce a peak tractive effort of 9,160 pounds. The total reaction per axle because of the dead weight of the entire arm assembly, with the model at the 50-foot radius, will amount to 22,800 pounds. Since such a high ratio of tractive force to normal wheel loading is not possible, an artificial means of wheel loading was necessary.

In this arrangement, keeper wheels, placed to bear on the underside of the horizontally supported track, are drawn up by means of tie rods passed through heavy coil springs. Spring compression will be adjusted until a total normal force of 61,000 pounds per steel-tired drive wheel is obtained. This wheel loading will permit maximum tractive effort without wheel slippage. Established practice is to keep facility tracks clean and free of oil to avoid lowering the coefficient of friction below 0.15. Fig. 3 shows essential features of the drive system and keeper wheels.

The selection of two 200-hp mill motors permits the use of low-voltage wiring and switchgear since the maximum direct voltage will be about 750. Two low-voltage generators connected in series and driven by a single 1,000-hp synchronous motor will supply the variable-voltage power to the armatures of the mill motors. The common connection of the two generators and of the two armatures will each be grounded, thus assuring low insulation stresses.

Excitation for the two generators will be provided by a d-c rotating-type amplifier rated at at least 30 kw and driven

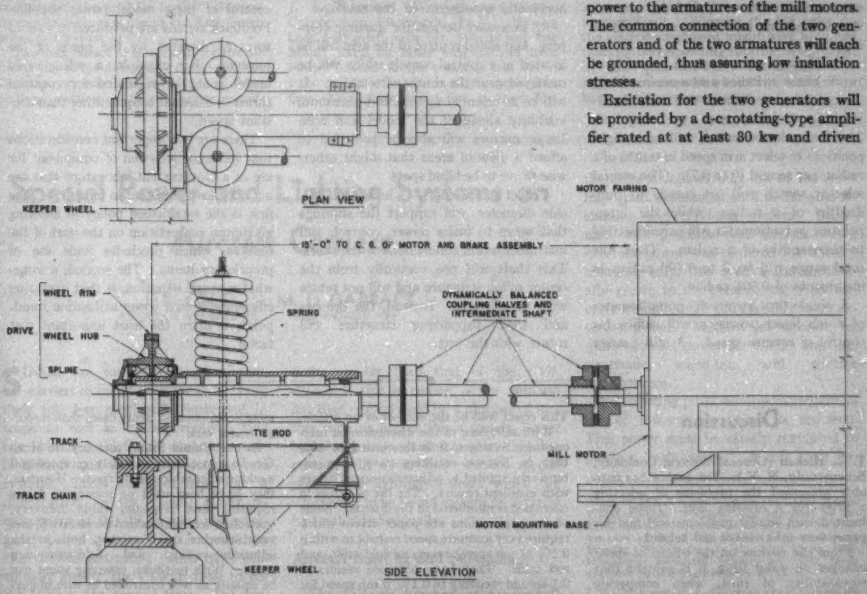


Fig. 3. Mill motor and drive assembly

by a 3-phase 440-volt motor. The control field of the rotating amplifier will be energized by means of a high-gain electronic amplifier, which in turn is fed the difference voltage occurring between the reference potentiometer and the output of the tachometer generator.

There are several unique ideas and provisions embodied in the contract specifications. Whereas it is common practice directly to couple speed-regulating tachometer generators to the controlled drive motors, a separate 12-inch idler wheel will drive the tachometer generator for this facility. Doing so will close the control loop around variables that are otherwise excluded. Wheel creepage and possible torsional oscillations of the wheel drive shaft, for example, are now included in the loop.

Although it is a fairly well established practice to apply electric machinery with a 15% service factor, the possibility of error in estimating frictional and windage loads and the likelihood of sizeable additional features being added to such research facilities make it advisable to use a somewhat larger factor. The specifications, therefore, require that all electric machinery be 25% larger than the minimum calculated size.

Several speed-controlling potentiometers, with a selector switch to connect the desired one into the system, will be supplied.

For high-speed control, an assembly of two selector switches and associated sets of precision resistors, and one interpolating potentiometer are required. The first selector switch will have eight positions to select arm speed in tenths of a radian per second (0 to 0.7). The second selector switch will set speeds to hundredths of a radian, while the interpolating potentiometer will permit setting to thousandths of a radian. Thus, the total range will be 0 to 0.799 radian in increments of 0.001 radian.

A single-turn reversing potentiometer of much lesser precision will suffice for control of reverse speed. A still coarser potentiometer will control inching and spot positioning of the arm.

In order to afford maximum flexibility of operation of the arm, separate acceleration and deceleration current-limit potentiometers are required. Each of these will be calibrated in 5% steps over a range of 30 to 100% of maximum permissible current. A special low-range current limit will also be provided to allow control from 10 to 34% of maximum.

Although the drive motors are not very costly, damage to them or to the facility as a result of malfunction could cost a great deal in terms of loss of testing time of a very important facility; hence a number of protective relays and safety devices are justified. The list includes slip detector relays, overvoltage relays, a zero voltage device, loss-of-field relays, instantaneous loop overcurrent relays, drive-motor thermal devices, ground detection relays, and centrifugal overspeed switches.

To avoid accidents to persons boarding or leaving the arm or to equipment being attached to the arm, a d-c loop contactor will open both sides of the loop as the arm speed approaches zero in the stopping sequence. As a further precaution, a mechanical holding brake mounted on the mill motors will aid in preventing accidental movement of the carriage.

All necessary devices for starting, stopping, and speed control of the arm will be located in a control console which will be mounted near the center of rotation. It will be so oriented as to afford maximum visibility ahead of the model and arm. Large mirrors will also be provided to afford a view of areas that might otherwise prove to be blind spots.

A steel hollow shaft, 14 inches in outside diameter, will support the sliprings that serve to make power, control, and instrumentation connections to the shore. This shaft will rise vertically from the center pylon structure and will not rotate with the carriage. Instead, the brushes and their supporting structure will rotate with the arm.

The slipring housing and the shaft will be compartmentalized into an "above-600-volt" section containing nine sliprings, each rated 250 amperes continuously; a "below-600-volt" section containing 10 sliprings, each rated 200 amperes continuously; a control compartment containing 59 sliprings arranged in nine sets of four unshielded, six sets of three each for 2-conductor shielded circuits, and one set of five rings for a 4-conductor shielded circuit. The instrumentation compartment will also have 59 sliprings. These, however, will be arranged in 13 sets of three and 4 sets of five for 2- and 4-conductor shielded circuits, respectively. Each instrument and control ring will be contacted by four silver graphite brushes for low-noise performance. The power brushes will be of copper graphite and will be arranged two per holder with two holders, 180 degrees apart, for each ring.

Another interesting feature of the rotating arm is its model power supplies. Wiring and switching will be arranged and equipment provided so that two 30-kw d-c generators can be used to power auxiliary testing devices or self-powered models. Error regulators and reference power supplies will be available and provision will be made for the independent control of these model power supplies. Feedback signals are produced in various ways as dictated by the needs of the moment. For example, a self-powered model could be regulated for constant thrust or constant torque rather than constant speed.

There are two important considerations that influence selection of equipment for use in a Government laboratory that are not necessarily existent in industry. The first is the established policy of allowing maximum competition on the part of the bidders, which precludes wide use of proprietary items. The second, a somewhat unusual situation, is that operating efficiency, from a power utilization standpoint, is often the least important cost factor.

Discussion

T. L. Hicken (General Electric Company, Schenectady, N. Y.): The author has capably presented the problems of selecting a drive for a rotating arm. Those confronted with similar problems will find his paper very informative and helpful.

From the treatise on the effects of speed changes on water drag, it is evident that repeatability of runs, when comparing hull design on ship models, plays a very important part in system performance. This could well be the subject of an article.

With reference to the discussion on mercury-arc rectifiers, it is interesting to note that in Europe rectifiers have been extensively applied to adjustable-speed drives with excellent results. The list of drives in operation is numbered in the dozens. Some of the applications are paper drives which require very accurate speed control to within 0.1% of top speed, reversing cold mills, and rod mills. They have obtained results of 0.1-second recovery to 0.1% of top speed for a 100% impact load. In Europe mercury-arc rectifiers have proved to be reliable and more economical in many cases than motor-generator sets.

In the United States mercury-arc rectifiers have not been as widely accepted and used for adjustable-speed drives. However, they have been used on numerous drives requiring fast response, rapid recovery, controlled acceleration and deceleration, and very accurate speed control, both in the adjustable-voltage and field-weakening range. With rectifiers, creeping speed can be equally as well controlled as with motor-generator sets. Rectifiers have proved to be both reliable and economical.

Within the next few years, we will see a greater use and acceptance of rectifiers in place of motor-generator sets for adjustable-speed drives. It is recognized that, due to a lack of experience on the part of electrical manufacturers in the use of mercury-arc rectifiers for adjustable-speed drives, the specifying of rectifiers would be a proprietary item, as expressed by Mr. Prucha at the end of his paper. This could be circumvented by specifying performance requirements, leaving the door open to the selection of equipment.

W. J. Roa, Jr., and **A. A. Godar** (Sverdrup & Parcel, Inc., St. Louis Mo.): The author's description of this unique facility should interest a diversified group of readers. The comparisons of various alternative drive arrangements are informative and the author and his associates should be complimented for the thoroughness of their preliminary investigations.

As is indicated, the design of a test facility such as the rotating arm involves considerations somewhat akin to prototype development for a complex military aircraft. In the former case, however, the final production run is one unit. Many conflicting factors must be resolved; the area of acceptable compromise is small, and at times elusive. The author's comments concerning the drive system illustrate this; for example, more power is desirable for acceleration but the effect of increased inertia absorbs nearly all of the increased drive power. Even after thorough preliminary design studies have produced logical choices and criteria, the need for further comparative analysis persists during the final design phase of a test facility.

The basic choice of equipment for the d-c power source involves many considerations. Other factors may be of interest in addition to those summarized by the author regarding comparison of rotating equipment and mercury-arc rectifiers. The principal drive system objectives include utmost reliability, rapid acceleration, quick attainment of preset speed within close tolerance, extremely precise speed regulation, means for speed reversal, and operation over a wide speed range. Considerations favoring mercury-arc rectifiers are small time constant and excellent transient response in the accelerating direction, less space requirement, and possibly lower first cost. Characteristics weighing against rectifiers for this facility are difficulty of assuring comparable reliability over the 25-to-1 speed range without the complication of including tapped primary transformers, lack of inherent ability to dissipate regenerative energy, need for preheating and maintaining temperatures suitable for intermittent operation, and increased arc-back probability with large angles of phase retard at the lower speeds. As Mr. Prucha implies, much effort and time precedes a series of tests. Highest probability of a successful first test run is needed for efficient operations. A second run following an arc back would have to be delayed due to disturbed water conditions. For this application, weighing of all factors led to the choice of rotating equipment for the d-c power supply.

Papers are seldom presented after a test facility has been operating for some time. It is suggested that the author present data on drive system performance, speed regulation and drift, and so on, after the facility is in operation.

J. A. Wasmund (Westinghouse Electric Corporation, Pittsburgh, Pa.): Having been associated with Mr. Prucha and the Model Basin on other projects and to a limited extent on the rotating arm drive, I am deeply interested in this paper. The first thing that strikes me is that Mr. Prucha, in a matter-of-fact way, has made an extremely complex and precision accuracy job sound almost easy. Laboratory accuracies of speed regulation of 0.1% to be obtained in a stabilizing period of 1/4 revolution after accelerating a 130-foot arm to a speed of 30 knots in 1/4 revolution by means of motor-driven, preloaded steel wheels on a steel track, whose waviness cannot be more than 0.001 inch in 6 inches and have a track deflection of not more than 0.01 inch, is no simple engineering feat.

A great deal of engineering effort has gone into study and elimination of various possible drive systems in order to select the optimum. Prior to reading this paper I have often wondered if the rotating arm drive should not have been driven by a single center-drive motor. Mr. Prucha has convinced me that this is not the case.

The use of a Ward Leonard system with two mill motors certainly seems to be the correct selection for the rotating arm drive and I extend my congratulations to Mr. Prucha for a very interesting and informative paper. My sincere hope is that, when the drive is put into service, my mental reservations regarding meeting certain portions of the specifications will prove to be groundless.

H. E. Prucha: If performance rather than fundamental equipment were specified as suggested by Mr. Hicken in order to leave the door open for either mercury-arc rectifiers or d-c generator sets to be furnished at a bidder's option, the simultaneous planning of space allocation and the management of installation contracts becomes more difficult. Whereas the bid price on the electric equipment might have been less than for corresponding rotating equipment, the total final cost to the Government is of prime concern, and whether it would have been any less is questionable.

Special Power and Lighting Systems on Aircraft Carriers

J. R. COLE
ASSOCIATE MEMBER AIEE

M. L. GARBACZ
NONMEMBER AIEE

SINCE the introduction of aircraft carriers to the Fleet of the U. S. Navy their role has continually increased in scope as well as in importance. This has been due chiefly to advances made in aircraft and associated weapons and systems. These advances have had a direct effect on the electric power requirements of the ship service and emergency power system. Fig. 1 shows how ship service generating plants have continually grown, the *CV4* in 1930 to the *CVA(N)* at the present time. The generating capacity of the *CVA(N)* is ten times greater than that of the *CV4*. Although this increased power demand has been necessitated by a number of factors such as higher ship's speeds, larger hulls, and more electronic gear, this discussion will cover only certain special power and lighting systems which are considered to be broad enough in scope to be treated individually.

Aircraft Starting and Servicing

The primary function of an aircraft carrier is to support and deploy aircraft, therefore it follows that one of the most important systems aboard is the aircraft starting and servicing system. Beginning with propellor-type aircraft where external power was required for battery charging only, this system has continually grown to the present era of jet aircraft which require considerably more external power. Modern jet aircraft require external power for engine starting, electronic warm-up, and routine maintenance.

For starting purposes approximately 33 kw of low-voltage power is required. This power must be closely regulated to produce "soft starts." A soft start im-

Paper 58-567, recommended by the AIEE Marine Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Washington, D. C., April 28-30, 1958. Manuscript submitted January 20, 1958; made available for printing February 20, 1958.

J. R. COLE and M. L. GARBACZ are both with the U. S. Navy Department, Bureau of Ships, Washington, D. C.

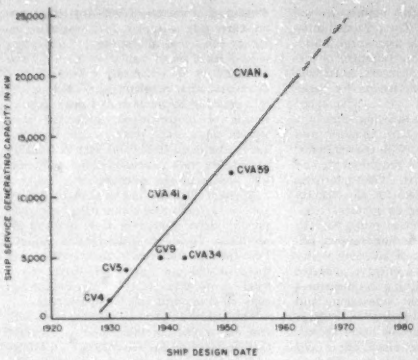


Fig. 1. Trend of plant growth for aircraft carriers

plies that the current to the starter motor does not exceed 1,100 amperes at any time. Such control prevents the development of excessive torques that would damage the motor or the mechanical coupling between the motor and the jet engine.

Servicing power is used for at least three different purposes:

1. It is used at combat-readiness stations to supply continuous power to interceptor aircraft to keep the aircraft's electronic gear and other vital equipment in a warmed-up "ready" condition. Up to 550 amperes of 28-volt d-c power and up to 60 kva of 115/200-volt 3-phase 4-wire 400-cycle power, depending on the type of aircraft, can be anticipated for this purpose.

2. It is required at almost any location on the flight deck for routine warm-up and check-out of the aircraft's gear just before the take-off.

3. It is also required on the flight and hangar decks and in aviation shops for routine test and maintenance of various pieces of aircraft equipment.

Aircraft starting and servicing systems vary with different classes of carriers. The most extensive system has been installed on the *Forrestal*-class carriers. These ships have a basic 400-cycle plant for this purpose supplied by two 600-kw 400-cycle turbogenerators. In addition, two 100-kw motor-generators are installed to supply starting and servicing power during periods other than launch when great blocks of power are demanded. Fig. 2 shows the general layout of the system. The 400-cycle power is generated by the turbogenerators at 1,000 volts and distributed to power panels located on the gallery deck. The use of this higher distribution voltage has resulted in a weight saving of approximately 20 tons. From the 1,000-volt distribution panels power is fed through banks of 1,000/120-volt transformers to provide 400-cycle power at 200/115 volts, 3 phase, 4 wire, for aircraft servicing. This power is supplied to the aircraft by means of portable cables which connect between receptacle outlets, located on the edge of the flight deck or recessed into the flight deck, and the receptable connectors on the aircraft itself.

Also, 1,000-volt power is supplied from the distribution panels through special rectifiers to provide d-c starting and servicing power. These rectifiers are designed normally to furnish closely regulated 28-volt d-c power for servicing and to furnish constant-current power for starting purposes on demand. The rectifiers are installed throughout the gallery deck and power from them is transmitted to the aircraft by means of portable cables and receptacle outlets strategically located to give the best coverage.

Fig. 3 shows a basic wiring circuit for distributing power from a rectifier power supply to an aircraft. To follow the circuit operation of Fig. 3 it is necessary to understand the control circuitry.

Terminal *J*, located in the low-voltage d-c power-supply control panel, is energized when the pilot pushes his jet-engine start switch or when the emergency jet start switch on the starting receptacle is closed. As soon as terminal *J* is energized, the constant-current relay coil is actuated which sets up a relay sequence in which power to the servicing receptacle is cut off and all the power from the power supply is applied to the $\pm CC$ jet starting terminal. This power is regulated with respect to current by means of magnetic-amplifier control circuits which limit the current to 1,100 amperes.

It has been determined that the most efficient electric starting of a jet engine occurs when the current to the starting motor is limited by current regulation at from 1,000 to 1,100 amperes with a finishing voltage of approximately 33 volts at the aircraft's starting receptacle. At the instant the starter motor is energized, its impedance is very low, causing its terminal voltage with 1,000 amperes flowing also to be low. As the starter motor gains speed, its back electromotive force increases, tending to decrease the starter current, but the power-supply current regulator will not allow the current to decrease. Therefore, the terminal voltage of the starter motor must increase

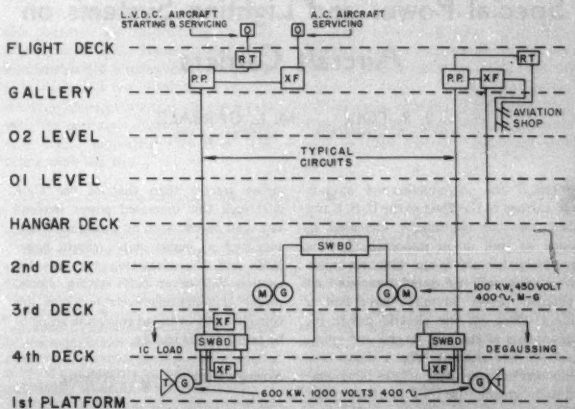


Fig. 2. The 1,000-volt 400-cycle aircraft starting and servicing system on the "USS Forrestal"

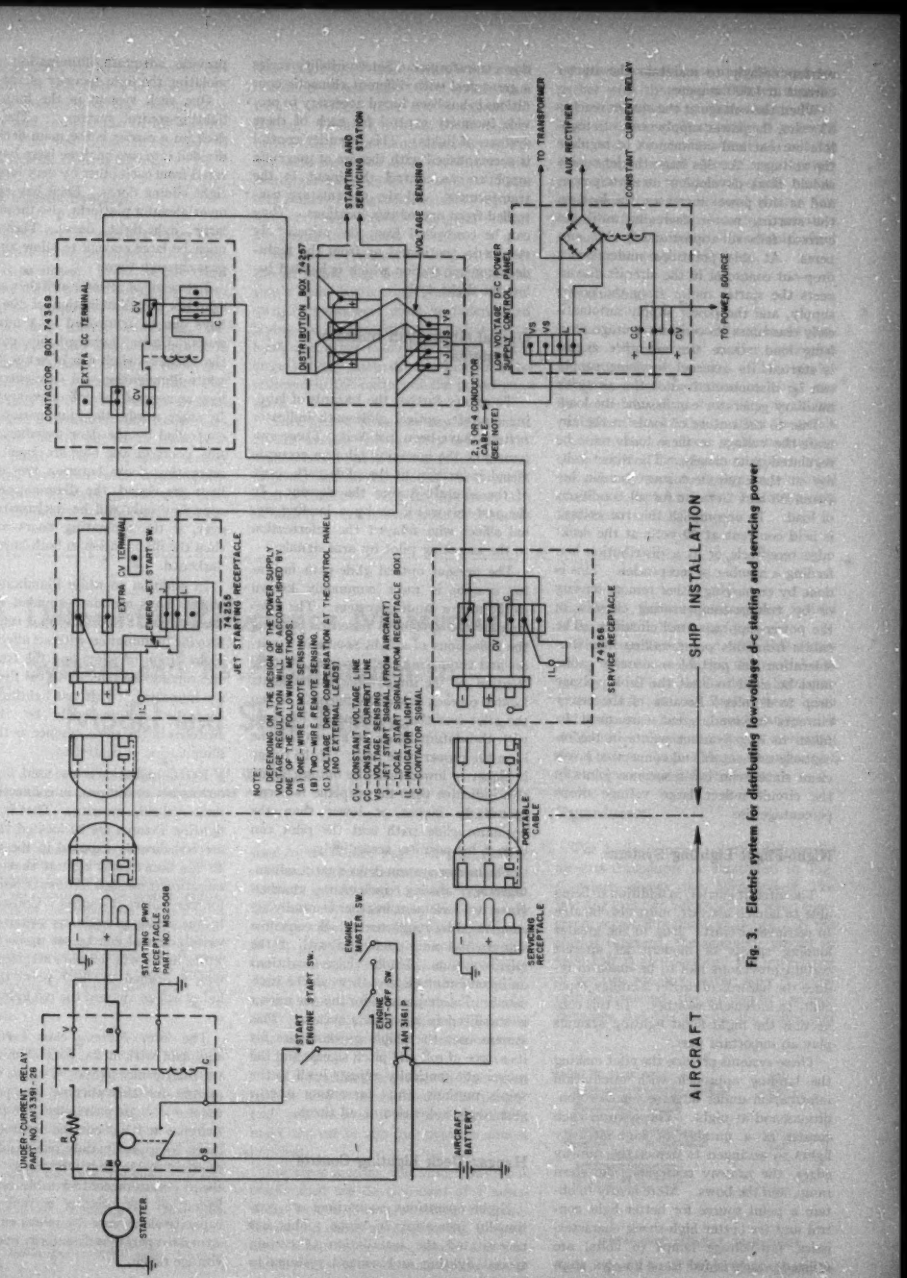


Fig. 3. Electric system for distributing low-voltage d-c starting and servicing power

correspondingly to maintain the starter current at 1,000 amperes.

When the voltage at the starter reaches 33 volts, the power supply ceases to regulate current and commences to regulate the voltage. At this stage the jet engine should start developing its own power, and as this power increases, the load on the starting motor decreases until the current falls to approximately 235 amperes. At this point an undercurrent drop-out contactor in the aircraft disconnects the starter motor from the power supply, and the power supply automatically reassumes its constant-voltage servicing load. Once the aircraft's engine is started, its external servicing supply can be disconnected since the aircraft's auxiliary generator can assume the load.

Due to the nature of loads in the aircraft the voltage to these loads must be regulated quite closely. The direct voltage at the equipment must remain between 25 and 29 volts for all conditions of load. To accomplish this the voltage is held constant at 29 volts at the deck-edge receptacle or at a distribution box feeding a number of receptacles. This is done by employing either remote sensing or by voltage-compensating circuits in the power-supply control circuitry. The cables from this point, taking into consideration the portable connector cable, must be sized to limit the cable voltage drop to 4 volts. Because of the heavy currents involved, great care must be taken to keep contact points in the receptacle connectors and connection boxes clean since even low-resistance joints in the circuit reflect large voltage drops percentagewise.

Night-Flight Lighting Systems

The aircraft carrier, in addition to being able to launch aircraft, must also be able to retrieve aircraft. Due to the greater landing speeds of modern jet aircraft certain provisions had to be made to reduce the hazards of carrier landings, especially in inclement weather. In this connection the night-flight lighting systems play an important part.

These systems provide the pilot making the landing approach with orientation information under adverse weather conditions and at night. The systems each consist of a number of high-intensity lights so arranged to depict the runway edges, the runway centerline, the stern ramp, and the bows. More nearly to obtain a point source for better light control and for better high-shock characteristics, low-voltage lamps (6 volts) are utilized; each is fed from its own step-down transformer. Since visibility varies a great deal with different climactic conditions it has been found necessary to provide intensity control for each of these systems of lights. This intensity control is accomplished with the use of magnetic amplifiers to control the input to the transformers. All the systems are controlled from one of two locations. They can be controlled from the primary fly station in the island or from the flight-deck control station which is located below the flight deck.

Optical Glide-Path Indicator Systems

To reduce further the hazards of landing aircraft, optical glide-path indicator systems have been installed. These systems give the incoming pilot an accurate visual indication of the glide-path angle of the aircraft during the descent. In the past this was done by a landing signal officer who relayed the information to the incoming pilot by arm signals.

The present optical glide-path indicator system is more commonly known as the mirror landing system. The principle of the system is based on viewing the reflections of a light source in a mirror and comparing the apparent vertical location of the image with datum lights located on each side of the mirror. When the pilot sees that the image is lined up with the datum lights he knows that he is on the proper glide path. If the image is higher or lower than the datum lights, this indicates that the airplane is correspondingly higher or lower than the optimum glide path and the pilot can correct his position accordingly.

The mirror system is used most advantageously during inclement weather. However, inclement weather is usually accompanied by rough seas which cause the ship to roll and pitch and with it the mirror system. Under these conditions an inconsistent glide path would be indicated. To compensate for this the mirror is stabilized in a one-axis system. This system uses the ship's gyrocompass for its source of roll and pitch signals and the mirror automatically adjusts itself to the ship's motions, thus presenting a true glide-angle indication at all times.

Hangar-Deck Lighting Control

Night operations on carriers are continually increasing in scope. This has necessitated the installation of certain special lighting and control systems to provide adequate illumination without violating the light security of the ship.

One such system is the hangar-deck lighting-control system. The hangar deck on a carrier is the main deck and is divided into two or three large bays separated from each other by very large light-tight sliding doors. Each bay opens up on an elevator platform, also through very large light-tight doors. These doors must be large enough to allow aircraft to pass through them.

By means of selector switches and door switches the illumination of the hangar bays can be controlled in a number of ways. Under darkened-ship conditions the selector switch can be set so that the white illumination will stay energized as long as none of the side doors are opened. In other words, the illumination is then controlled by the door switches. If the side doors in one bay are open but the separating doors between two adjacent bays are closed, the illumination in the open bay only will be darkened. However, if the separating doors are open then the illumination in both bays will be darkened.

In addition to white illumination, the hangar bays are also provided with red illumination. Red lighting is required to provide illumination without affecting the night vision of pilots and the crews that accompany the aircraft to the flight deck for launching. This red lighting can be controlled automatically by the door switches in the same manner as the white illumination is controlled.

Red illumination is also used, depending on certain conditions, in connection with darkened-ship security. That is, the red lighting fixtures are so located that they are not directly exposed to the weather. It has been found by test that the illumination from such fixtures is visible only for very short distances. Consequently, by the setting of a selector switch the door switch control can be set up so that the white lights will remain energized until a door is opened at which point the white lights will go out and the red lights will go on.

The later *Forrestal*-class carriers are equipped with a d-c control system for the hangar-bay lights. The use of direct current for this purpose has permitted the use of single-pole in lieu of 3-pole door switches and has cut the control cabling down to approximately one third of that required in a-c control systems. Also, in the d-c system control cable can be selected on the basis of current-carrying capacity alone since d-c relays can be designed to operate satisfactorily over a wide voltage range.

To supplement the lighting control circuit a special test circuit is required to check the condition of the various door switches and components of the control system. This test circuit is essential to insure a reliable control circuit.

Guided Missile Systems

The latest aircraft carriers under construction are being provided with guided missile capabilities. From an electrical viewpoint this involves supplying power for missile guidance systems, missile check-out, and missile launching. Although this appears simple enough it is not actually so, since most of these systems require very nearly transient-free power closely regulated in voltage and frequency and at voltages and frequencies other than that of the ship's basic generating plant. For example, certain equipment associated with missile guidance requires 400-cycle power with voltage and frequency tolerances not to exceed $\pm 1/2\%$ of steady-state values and transients of less than 1% of steady state with a recovery time not to exceed 0.1 second. In order to meet these requirements it has been found necessary to isolate each such load and provide a special motor-generator set for its supply. These sets employ a magnetic clutch to regulate frequency and specially designed voltage regulators to regulate voltage.

The motor-generators, along with other special power supplies, are fed from the ship's basic distributing system. Consequently, great care must be taken when laying out the system to isolate big motor loads and other loads that would create large voltage fluctuations which may be reflected in the output of the power supplies.

Since loss of power to any component of this system can be very serious, the system is laid out so that alternate sources of power can be provided in the event of such a condition. To do this most economically a very limited number of standby power supplies are installed and together with the required power supplies are so arranged to form secondary distribution systems. Eventually it is hoped that with the development of suitable line regulators and electric speed governors the large number of small motor-generators can be replaced with a few large centrally located motor-generators to supply the system. This would effect a saving in weight and space.

Conclusions

The number and variety of loads associated with aircraft carrier operation is continually increasing. This creates many new problems, most of which can be solved locally; however, there are problems that can and do affect the whole electric system. These problems must be given careful consideration with a view toward any major change in electrical plant philosophy required for future ship installations.

A Differential Reluctance Measuring System and Its Application to Model and Ship Testing

MEREDITH W. WILSON
NONMEMBER AIEE

Synopsis: A measuring system is described which has been developed at the David Taylor Model Basin for use in model and full-scale ship testing. The system utilizes transducers of the differential air-gap type, special designs of which permit measurement of torque or thrust in rotating shafts without the use of slip rings. The system includes instrumentation for direct recording and for visual or automatic digital readout. Accuracies of $\pm 1/4\%$ or better are obtainable.

THE TESTING of models and full-scale ships requires precise measurement of many quantities, such as torque, thrust, forces, and moments. The measuring system described herein is widely used by the Model Basin for such testing.

Since many of the measurements must be made in submerged models, it is essential that the transducers be completely waterproof. In rotating shafts, the use of slip rings and brushes would require very thorough shaft seals and gasketing. Aside from the waterproofing requirements, the maintenance and the unavoidable electrical noise resulting from slip ring assemblies would present a continuing problem.

For torque measurements in high-speed turbine shafts on shipboard, the peripheral velocities and ambient conditions make the use of slip-ring systems almost hopeless.

For these reasons, the Model Basin has emphasized the development of a measuring system through which measurements and observation of forces, torques, and thrusts in rotating shafts and machinery may be made without the use of slip rings. Such a measurement system is described in this paper.

The transducers used with this measuring system are of the differential air-gap type. They are made in three general forms: (1) a displacement gage (Magnigage) which, in combination with a flexure, can measure forces or moments; (2) a torque gage (Magnitorque); and (3) a thrust gage (Magnithrust).

Transducers

The basic principle of the differential air-gap transducer is illustrated in Fig. 1, which shows the Magnigage. The winding is in two halves which, when energized, set up flux paths as shown. If the core is centered, the magnetic circuit in each half is identical and the impedance of the two windings is the same. If the core is displaced, the air gap at one end increases and the gap at the other end

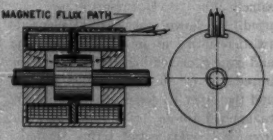


Fig. 1. David Taylor Model Basin Magnigage

Paper 58-548, recommended by the AIEE Marine Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Washington, D. C., April 28-30, 1958. Manuscript submitted January 17, 1958; made available for printing February 26, 1958.

MEREDITH W. WILSON is with the U. S. Navy Department, David Taylor Model Basin, Washington, D. C.

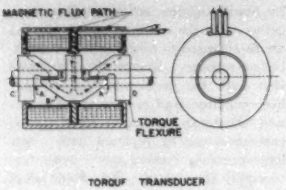


Fig. 2. David Taylor Model Basin Magnitorque

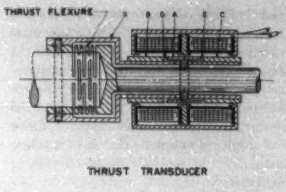


Fig. 3. David Taylor Model Basin Magnithrust

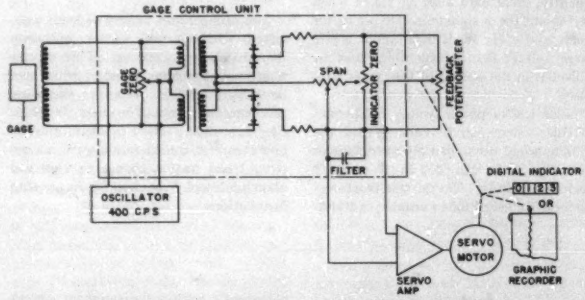


Fig. 4. Schematic circuit diagram, differential reluctance measuring system

decreases. The resulting differential impedance change alters the voltage drops across the windings and gives an output signal.

The torque transducer (Magnitorque) is shown in Fig. 2. Two necked-down sections of the shaft, A and A,[1] having the same diameter and length, form the flexural elements. Three rings made of magnetic material are fastened to the shaft over these flexural elements; one ring, B, is fastened to the land between the flexure sections, and the other two rings, C and D, are fastened to the shaft at the ends of the flexures. The rings are made so that a series of air gaps are formed between the center ring and the two end rings.

With no torque applied to the shaft, the active gaps E and F are equal and the impedance of each winding is the same. When a clockwise torque is applied, as shown, the flexural elements twist, resulting in air gaps F increasing and gaps E decreasing. This causes the impedance of one winding to increase and the other to decrease proportionally to applied torque. The shaft is made from a nonmagnetic spring material as beryllium copper or K-monel to avoid short-circuiting the magnetic flux.

The thrust transducer (Magnithrust) is shown in Fig. 3. The flexural element consists of an annular section of the shaft in which a series of slots is cut perpendicular to the shaft, forming a type of compression spring. The shaft extension on one side of the flexure carries a ring A. This ring is centered between two similar rings B and C which are secured to sleeve S, which is fastened to the shaft on the opposite side of the flexure.

With no thrust applied to the shaft, the active air gaps D and E are equal and the impedance of each winding is the same. When a positive thrust is applied as shown, the spring is compressed, thereby moving ring A to the left. This movement decreases air gaps D and increases gap E, resulting in a differential impedance change and a corresponding output voltage proportional to load.

Circuitry and Instrumentation

For measuring the outputs of these transducers, an automatic null balancing system is used, in which the transducer and the indicator or recorder form a closed-loop servosystem. The transducer output is balanced by a potentiometer. Any error signal from the transducer, resulting from torque, thrust, or other force, is amplified and drives a servomotor which positions the potentiometers to restore electrical balance, or null, to the system. The amount that the potentiometer moves is a measure of the force.

Fig. 4 shows a simplified diagram of the circuitry used. It will be noted that the gage signal is converted to direct current before balancing against the feedback potentiometer. This eliminates the need for phasing adjustment. Controls are provided for span adjustment, for gage zeroing, and for filtering or smoothing the signal output. Silicon diodes provide noise-free stable rectification of the reference and signal voltages.

Readout is provided in two forms: as a graphic record on a wide-scale recorder, or as a number on a digital indicator.

The graphic recorder is a servo-balancing-type instrument with a 10-

Fig. 5. Graphic recorder assembly

Fig. 6. Digital indicator Taylor Model Basin type 299

Fig. 8. Propeller dynamometer assembled

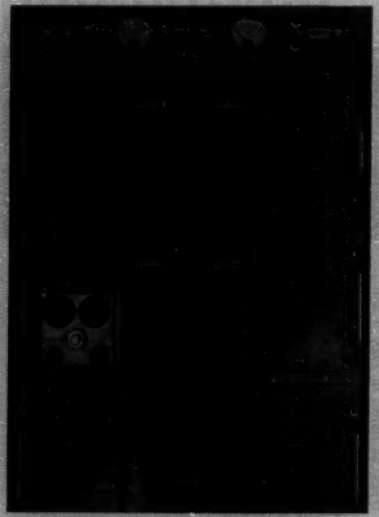


Fig. 7. Digital indicator Taylor Model Basin type 301

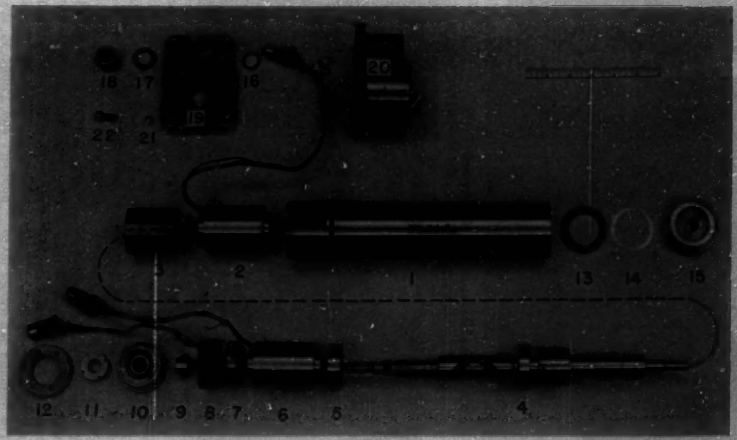


Fig. 9. Propeller dynamometer exploded

Fig. 10. Contra rotating torpedo dynamometer

Fig. 13. Turbine torquemeter

Fig. 11. Thrust transient dynamometer

Fig. 14. Turbine torquemeter armature

inch-wide scale. Response time may be 1 second or 1/4 second for full scale, depending upon the particular recorder used. A complete 2-channel system includes a 2-channel recorder, a 2-channel control unit, and an oscillator. (See Fig. 5.) The oscillator can supply up to eight channels.

The digital indicator has been made in two types, as shown in Figs. 6 and 7. The former is a single-channel unit, consisting of servomotor, a 10-turn helical potentiometer, gear box, and a digital readout. Nominal full-scale reading is ±1,000 divisions, giving a reading resolution of 0.1%. Gear ratio is such that the maximum travel is ±1,050 divisions, giving some overtravel at each end.

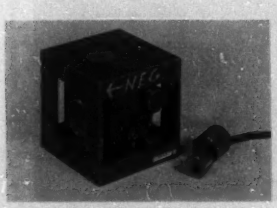


Fig. 12. Modular force gage

The digital readout is a visual-type rotary counter with a commutator on each wheel to provide for remote electrical readout. A 10-toothed wheel and a solenoid are provided also to prevent possibility of the counter stopping between numbers and failing to read out. To take a reading, the motor is stopped and a voltage pulse is applied to the solenoid. This pulls the counter in to the nearest whole digit.

This indicator uses the same gage-control unit, oscillator, and servo amplifier as are used with the graphic recorder.

The second type of digital indicator is a 2-channel unit with gage-control units, servo amplifiers, and servomotor assemblies contained in a single case (Fig. 7). The only other equipment required is the oscillator unit, which supplies up to eight channels. Readout at present is visual only. However, provision has been made for future addition of a digital converter of the command readout type.

For recording of transients or study of signals varying more rapidly than the servo-type graphic recorder will handle, a carrier-type amplifier is used, with either a string oscillograph or a high-speed direct-writing recorder of the ink pen or heated stylus type.

Applications

The most widespread application of this measuring system at the Model Basin has been in model propulsion dynamometers to measure torque and thrust in propeller shafts. These dynamometers as developed over the past few years have a number of desirable features:

1. They are small in size. The dynamometers are quite compact. Thus, they may be used in small models or mounted in a stern tube close to the propeller.

2. They need no slip rings. There are no mechanical connections or restraints between the shaft assembly and the housing, except for the bearings which center the shaft. This simplifies the mechanical design and manufacture, eliminates signal errors due to slip rings, conserves space, does away with frictional torque of slip rings, and with the frequent servicing which slip ring assemblies generally require.

3. They are submersible. The elimination of slip rings permits free-flooded operation of the dynamometers. The coils are vacuum-cast with an epoxy-type plastic material, and all connectors are watertight. All parts throughout are made of corrosion-resistant material.

4. They have high natural frequency. The sensitivity of the transducers is such that the flexural elements may be made quite stiff. Thus the dynamometers are

suitable for many types of transient torque and thrust measurements. As a general rule the transducers are designed so that the deflection at the air gap corresponding to full load is approximately 0.003 to 0.004 in.

5. They are of the transmission type. In contrast with the reaction type, these dynamometers are independent of the driving motors. Thus a motor may be replaced easily in case of trouble or if a motor of different characteristics is required. Also, it is possible to synchronize two or more shafts by gearing the input shafts together. This cannot be done with reaction-type dynamometers.

Figs. 8 and 9 show a typical propeller dynamometer of recent design. Outer diameter of the cylindrical portion is $2^{9}/_{16}$ in. and over-all length of the housing is $11^{1}/_{4}$ in. These dynamometers are made in two capacities: 50-lb (pound) thrust, 50-in.-lb torque; and 100-lb thrust, 100 in.-lb torque. The two are identical except for flexure design. In addition to the torque and thrust transducers, an electromagnetic revolution-counter pickup is included. This consists of a slotted disk and a pair of pickup coils which produce ton pulses per revolution.

In addition to these general-purpose propeller dynamometers, other dynamometers and devices have been designed to meet special requirements. Some examples will be given.

Contra-Rotating Torpedo Dynamometer

This dynamometer (Fig. 10) was designed to measure torque and thrust in the counter-rotating shafts of a full-scale torpedo. In this dynamometer, one torque and one thrust element are made hollow, so that the other shaft can pass through. Torque and thrust ratings are 3,300 lb-in. and 600 lb per shaft at 3,000 rpm.

Thrust Transient Dynamometer

The purpose of this dynamometer (Fig. 11) is to measure thrust vibratory forces in model propeller shafts. In order to make its frequency response as high as possible, the flexure was made very stiff (design load of 50 lb gives a deflection of about 0.008 in.). Also, to permit installation immediately forward of the propeller, the dynamometer was made quite small ($1^{1}/_{8}$ in. in diameter by $4^{1}/_{2}$ in. long). A Magnithrust-type gage is used. The natural frequency with a 5-lb mass attached is about 240 cycles per second.

Modular Force Gage

A number of force balances have been designed and built for measuring forces and moments in towed models. Each unit is in the form of a 4-in. cube (Fig. 12) and is sensitive to forces in one direction only. By use of connecting brackets and fixtures, these balances are used in combinations to measure lift, drag, side force, pitching moment, and yawing moment. Magnigage-type transducers are used, encapsulated to permit submerged operations. Capacities of 50 lb, 250 lb, and 500 lb have been built. Each is provided with limit stops which will permit loads to 1,000 lb or so, such as will occur when a large model is lifted from the water.

Turbine Torquemeter

This torquemeter (Figs. 13 and 14) was designed to measure torque in a high-speed turbine shaft. The high speed (8,000 rpm), high temperature (150 degrees Fahrenheit), and oil-mist-filled location make it almost impossible to use slip-ring systems. Magnitorque-type gage is used, with the coil assembly mounted on high-speed ball bearings. Capacity is approximately 17,000 horsepower.

Conclusions

The applications described herein represent only the particular needs of the Model Basin to date. Transducers with more sensitivity, higher frequency response, or different sizes and shapes can easily be made if required.

Instrumentation now in use is generally satisfactory but continuing efforts at improvement are under way. Improvements in components have considerably increased the stability and accuracy, and some consideration is being given to replacing the servo system with a direct voltage-to-digital converter.

A readout system has been designed and built to collect the data from several channels of such instrumentation and to transcribe these data simultaneously on an electric typewriter and a card punch. Other similar systems will probably be forthcoming.

Design Features of Large Semiconductor Rectifiers

I. K. DORTORT
MEMBER AIEE

PARAPHRASING Gertrude Stein, a rectifier is a rectifier is a rectifier. The many different service requirements, compounded by many economic factors, all influence the choice of design features available to the manufacturer. To obtain the best rectifier to fill specific requirements, one needs to make a careful evaluation of all the factors involved.

The purpose of this paper is to discuss briefly the characteristics of the important features of large semiconductor rectifiers, and how, from our viewpoint, they fit into the total picture. The author pleads guilty at the outset to the possible charge of prejudice. It is hoped, however, that this presentation of an engineering framework will provoke discussion of benefit to all concerned.

Some of the more important features are listed in the following. The sequence in which they are listed is not necessarily the order of their importance, which may differ from case to case:

1. Unit size and subdivision.
2. Transformer connection.
3. Cell PIV (peak inverse voltage) rating and voltage division.
4. Cell current ratings and current balancing.
5. Cooling.
6. Protection.
   a. Internal faults.
   b. External faults and overloads.
7. Switching and control.
8. Regulation.

Many more items could be added to this list and much time spent on the study of

Paper 58-516, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, East Lansing, Mich., May 5-7, 1958. Manuscript submitted January 13, 1958; made available for printing February 24, 1958.

I. K. Dortort is with the I-T-E Circuit Breaker Company, Philadelphia, Pa.

the interrelation of various features. It is hoped that the discussion which follows will be sufficiently clear to bring out these influences, at least in the more important cases.

Unit Size and Subdivision

As with other apparatus, the cost per kilowatt of a rectifier descreases as the unit rating increases. However, there are practical limits to the size of a unit, imposed by considerations of design and application, namely:

1. The fraction of the total capacity represented by one unit and the consequence of the loss of this fraction.
2. Practical limits of voltage and current division.
3. Availability of switchgear in the required ratings.
4. Design limitations of high-current transformer windings, which involve considerations of efficiency and regulation.
5. The necessity of phase multiplication with particular reference to reactance drop and telephone interference.

With reference to item 1, the larger the total number of units supplying a single load, the less important it becomes to maintain continuous service on any one unit. On the other hand, a process which can take interruption without harmful effect, other than loss of production, may be supplied by a single large rectifier. In many cases, a satisfactory compromise between cost and service continuity is provided by subdividing the rectifier into partially independent sections.

With reference to item 2, regardless of the means used to assure current and voltage division among parallel- and series-connected cells respectively, there is for each system a point of diminishing returns. The rapidly increasing voltage ratings of commercially available diodes assures one that the number of series-connected cells required for high-power industrial rectifiers will seldom be a serious problem. There is available today satisfactory means for obtaining proper division of inverse voltage among as many series-connected cells as may be required for the highest direct voltages in common use.

This is not true with respect to the number of cells needed in parallel to achieve desired current ratings. Current-balancing reactors permit the maximum number of parallel-connected diodes without derating. However, even with the most carefully designed balancing reactors, the maximum unbalance is still a function of the number of cells connected in parallel. This unbalance can be reduced by designing reactors with very low magnetizing currents, generally at some cost of efficiency, greater space requirements, material, and labor. Improvement can also be obtained by using coded cells together with reactors, but an upper limit still exists.

With reference to item 3, it is doubtful that primary switchgear ratings will ever become a limitation unless the primary supply is of very low voltage. This is hardly likely to be the case in an installation of large rectifiers. Secondary and d-c gear, if required, will certainly be pushed to their maximum ratings. It is quite likely that this limitation will, in many instances, force a division of rectifiers into sections with corresponding division of transformer windings.

With reference to item 4, monocrystalline semiconductors, because of their low forward losses, are particularly useful in high-current low-voltage conversion units. Application requirements have already pushed transformer design to the limits of economy, specifically in conductor cross section, eddy current, and impedance, so that subdivision has become necessary.

Balanced transformer design for two or more secondaries per phase becomes increasingly difficult at the current rating increases, or the voltage decreases. This is equally true of double-Y-connected transformers although the possibility of reducing the number of directly paralleled cells by this means will often favor this design for low-voltage applications. By the same token, the triple-diametric connection offers still further reduction of paralleled cells in the very low voltage field. Both the double-Y and triple-diametric connections create higher crest voltages and are advantageous only when no series-connected diodes are required to withstand the higher voltage.

With reference to item 5, the maximum feasible phase number for a single rectifier transformer is 12. Even this number may be impracticable for heavy current secondaries, particularly at low voltage: the number of turns per secondary may be too small to permit balanced Y-delta secondaries. Zig-zag windings are still more difficult, even disregarding their cross connections.

Twelve-phase operation is not always sufficient to avoid excessive reactance drop or telephone interference. Where further phase multiplication is required, it will often be found more economical, all things considered, to build two or more small rectifier units, phase-shifted from each other, rather than one large unit with several phase-shifted secondaries.

Transformer Connections

Except for rectifier voltages higher than those generally encountered in high-capacity units, three basic circuits come into consideration for most industrial applications.

1. Six-phase double-way (3-phase bridge).
2. Six-phase double-Y (with interphase transformer).
3. Triple-diametric (with 3-phase interphase transformer).

In the order listed these three connections show, for a given rectifier rating, increasing transformer size, and increasing crest inverse voltages. However, in the same order, the number of cells directly connected in parallel decreases. If the maximum permissible crest voltage is not limited by the PIV ratings of a single cell, the 6-phase double-Y or triple-diametric transformer may prove advantageous despite the increased cost of the transformer plus the cost of the interphase transformer. The choice is not always easy: the 6-phase double-way connection should be given first consideration unless the application definitely calls for one of the other connections.

Using two 3-phase bridges, each connected to a Y and delta secondary of the same transformer, 12-phase operation is obtained. Furthermore, this arrangement inherently divides the rectifier unit into two sections which can be protected and switched independently to decrease the loss of capacity in case of trouble.

The 6-phase double-Y connection also provides subdivision of the unit, but it is not advisable to operate one half without the other for any lengrh of time because of the d-c saturation of the transformer core, and badly distorted primary currents.

The triple-diametric connection divides the rectifier unit into three sections. With a properly designed interphase transformer and by taking advantage of the inductance of the d-c busses in low-voltage applications, it is possible to operate with one section out of service.

Cell PIV Rating and Voltage Division

Because of the lack of knowledge of the factors that make semiconductor rectifiers fail on voltage, and because of the poor division of inverse voltage heretofore obtained across series-connected cells, a total PIV cell rating of three times the circuit crest voltage has been advocated. This high cell rating is also intended to provide a margin of safety against the high-frequency oscillations that take

place in many rectifiers at the end of each commutation period, and the unpredictable switching surges and other disturbances that occur on power systems. Depending on the nature of the load and the type of switchgear employed, switching surges may originate in the d-c as well as the a-c systems. Surges may also occur due to the operation of improperly designed or applied cell fuses. Nonlinear resistors in combination with capacitors have been hopefully but dubiously applied to reduce these effects.

Two schools of thought exist on the question of voltage division. One school holds that no special means of voltage division is required other than a certain degree of matching of cells. This school argues that the cells best able to take the higher voltage are the ones with lowest inverse leakage current, and they get the highest voltage automatically. The author's experience does not bear this out.

Among those who believe that some form of voltage division is required, three systems have been employed:

1. In the early days it was common practice to use a separate transformer, or at least a separate transformer winding for each of the series cells so that accurate voltage division was assured. To a large extent this nullified the economy of large units, since in effect, there were two or more complete rectifiers of individually low voltage ratings connected in series to obtain the required output.

2. To regain the advantage of large units, more recent practice has been to connect voltage-divider resistors across the individual cells of a series string. For reasonably accurate voltage division by this means, the current in the divider resistors must be greater than the difference of inverse leakage currents of the worst and best cells. This scheme could produce excessively high losses and seriously impair the efficiency of an inherently efficient device. By careful coding of the cells according to their inverse current, it is possible to reduce the required resistor current and losses.

3. A method using relatively small and inexpensive voltage-divider transformers has been developed and is in successful operation. The economy of full-voltage transformer windings is retained and absolutely rigid voltage division is obtained without added losses. This system has the additional advantage of providing ready means for detecting a single bad cell in a series string and taking corrective action without first inducing failure in the remaining series cells. A reduction of the ratio of cell PIV rating to circuit crest voltage is indicated and will undoubtedly occur as experience is gained.

Cell Current Ratings and Current Balancing

Current ratings of individual cells are increasing rapidly and no one can say what ratings will ultimately be achieved. Nevertheless, it is likely that regardless of this increase, there will always be a need for operating a number of cells in parallel. With no provision made for forced current balancing, derating factors for parallel cells must be applied, even with carefully coded cells.

For large, high-power rectifiers, no one should seriously consider individual series resistors to force proper distribution of current. A far better means is the use of current-balancing reactors. If well designed, the first cost of these reactors will be offset by the reduction in the number of cells required to carry a given current, and will not add appreciably to the cell and copper losses. Additional savings are realized in the longer life of cells.

The number of rectifier cells that can be properly balanced in a single bank will be an important factor in determining the maximum rating of a rectifier section and of the total unit.

Cooling

The cooling of semiconductor rectifiers deserves the utmost attention. Not only can their life be seriously shortened by overtemperature, but the danger is greatly magnified by their very small inherent thermal capacity. It is important, then, to provide additional thermal capacity as close to the junction as possible and also to obtain the lowest operating temperature consistent with first cost and operating costs.

The first approach to cooling was from the direction of high-velocity air blast. There will always be some applications in which this is the logical approach. In the usual type of large industrial installations this method presents serious difficulties. Adequate cooling must be provided throughout the entire year and, therefore, the maximum ambient temperature becomes the first limitation for direct air cooling. In many installations the air is badly contaminated and must be washed, or at least filtered. For both these reasons some installations have been designed with closed air-recirculation systems in which the air is cooled either by air conditioners or air-to-water heat exchangers. Because of the inherently high temperature drop at air-cooled metal surfaces, high air velocities are required and this in turn means large volumes of air must be handled at comparatively high pressures. Large blower motors decrease the over-all efficiency of the rectifier, and their energy consumption is added to the amount of heat to be dissipated.

The liquid-cooling system is certainly a more direct approach. It provides much better heat transfer at the primary cooling surfaces, and eliminates massive extended surfaces, air-handling equipment, and ducts. Heat is transferred to the ultimate sink by the most direct and efficient way, resulting in the lowest temperature difference. It permits that most compact arrangements with the lowest expenditure of energy and first cost. A large saving in building costs can often be realized.

Two factors have probably contributed to retard the general adoption of liquid cooling:

1. The unhappy recollection of corrosion and electrolysis problems with mercury-arc rectifiers.

2. The necessity in some designs of breaking into the cooling system to change cells, with all the disadvantages and possibilities of leaks ever present in such systems.

In order to avoid difficulties with electrolysis and corrosion, various liquids have been suggested for the coolant. The liquid chosen must, of course, have a high resistivity and must not attack materials with which it comes in contact or be contaminated by them. Deterioration of the liquid, if any, must be very slow and the resulting reaction products must not be harmful. Another important consideration is flammability. If the coolant can burn, the circulating pumps may feed fuel into a fire resulting from a leak or break.

Water is, of course, one of the best cooling media known. All of its bad properties are caused by bad associates, which it picks up most readily. In the past, it has been used with anticorrosion agents such as sodium dichromate; sodium dichromate lowers its resistivity and "targets" are used to localize the resulting electrolysis.

It is certainly more logical to remove or neutralize continuously those substances in the water which are harmful. Today, it is possible to do so by passing the water through ion exchanges, which, when properly applied, easily maintain the water at a resistivity in excess of 12 megohm-centimeters. Laboratory tests and early returns on a unit so equipped, give confidence that the most vulnerable parts of the cooling system, the terminal fittings, have a minimum life expectancy of 20 years.

Although it is undoubtedly true that the most efficient liquid cooling is obtained where the case of the semiconductor cell has integral extended surfaces in direct contact with the cooling liquid, a very satisfactory structure has

been designed in which the case is cooled by contact with a cold bus. This permits a tremendous reduction in the number of gasketed joints, reducing the number of possible trouble spots and at the same time increasing the ease with which defective cells can be replaced.

Protection

Internal Faults

When a single cell is used for the full voltage, i.e., when the system does not require series-connected cells, a single high-speed current-limiting fuse per cell offers excellent protection. Fuses of this type have been developed specifically for the protection of semiconductor diodes with special attention given to the problem of limiting recovery voltages.

When two or more cells are connected in series, no fault current is available to blow a fuse until the last series cell has failed. The failure of the first cell imposes higher inverse voltages across the remaining cells so that a bad cell will eventually cause the failure of one or more good cells before the fuse can operate. If each of the series elements consists of several parallel-connected cells, the failure of a cell in one bank produces overvoltages on all cells in the other banks. It has, therefore, been common practice to have no parallel cells in a series string, but to parallel as many strings as required by the rating of the unit. This practice increases the size and cost of the rectifier "stacks" and greatly increases the difficulties of liquid cooling.

To permit common bussing and cooling of all parallel-connected cells, it is necessary to detect a single bad cell and remove as quickly as possible the resulting overvoltage on the remaining cells connected in series with it. The voltage-divider transformers mentioned previously provide a ready means for detecting a bad cell and supplying a strong actuating signal to a high-speed protective device.

This device, in one, form short-circuits the a-c bus of the rectifier section involved. It is not necessary to apply a full short circuit to the rectifier transformer, but only sufficient overload to pull the voltage applied to the cells down to a level below d-c bus voltage. The overload so imposed is then used to trip a secondary or primary circuit breaker and clear the fault.

A 10,000-ampere 300-volt rectifier equipped with voltage-divider transformers and short-circuiters has been in service since early November 1957. Only five cells have failed so far; four of them in the first week of testing. It is noteworthy that the system used for locating bad cells, by applying a low a-c test voltage after shutdown, has been found too insensitive. The speed of the protective equipment prevents the failed cell from going to complete breakdown, resulting in indefinite readings.

In the case mentioned, the rectifier unit was split into two 5,000-ampere sections arranged for individual shutdown of each section in case of cell failure. There is sufficient other capacity connected to the bus so that the loss of 5,000 amperes does not curtail production. It is perfectly feasible to increase the size of the voltage-divider transformer so that it can supply sufficient current to blow a cell fuse. When this is done, the PIV rating of the cells must be such that they will not be harmed by the full circuit voltage applied to the remaining good cells during the time required to blow the fuse.

External Faults and Overloads

Cell fuses do not afford protection against external faults and overloads. Because of their extremely small thermal capacity, high-speed protection is required, even for overloads of less than fault magnitude. Protection may be provided by high-speed bypass contactors or "short-circuiters" and breakers. In most chemical plants, reasonable care in the layout of busses and equipment can virtually eliminate the possibility of short circuits. However, the possibilities of faults in parallel-connected conversion equipment require the utmost care in the selection and co-ordination of protective equipment.

Again, it should be noted that proper division of current and voltage, together with adequate liquid cooling, reduce the danger of cell damage during the time it takes the protective equipment to function.

Switching and Switchgear

When the possibilities of external faults are so remote that the cost of expensive protective equipment is not considered justified, and particularly, if parallel-connected conversion equipment consists only of semiconductor rectifiers, switchgear may be required only for the purpose of connecting and disconnecting the rectifier. With sufficient range of voltage control available to reduce the load to zero, no-load switches are sufficient; otherwise, load-break switches on the a-c side, primary or secondary, will do the job economically.

Fused disconnects on the primary side of individual units, with perhaps a single main breaker for a group of rectifiers, avoid the need of many expensive high-voltage breakers. This arrangement will be found more generally applicable than in mercury-arc rectifier installations, with individual cell fuses replacing the protective function of the anode breaker.

Simple d-c switches can be used in most cases for isolating the rectifier from the bus. In many cases of isolated loads, even these are not needed. Large current-limiting fuses offer good backup protection against the remote chance of feedback from the d-c bus into an internal fault not properly cleared by the normal protection.

Secondary disconnects, or breakers, when needed, are used to isolate sections of a rectifier arranged for continuity of partial load service. Secondary switches offer an additional advantage in preventing surges incidental to primary switching from reaching the diodes.

Regulation

In many industrial applications, no control of the output rectifier is necessary other than the proper design of the rectifier transformer to provide the desired inherent regulation. The droop of the volt-ampere characteristic is, of course, influenced by the choice of connection and impedance. It would be possible to obtain flat compounding by special treatment of the interphase transformer in the manner sometimes used with mercury-arc rectifiers before the advent of phase control. This system increases the thermal loading of the cells and exhibits other serious disadvantages. It is not recommended.

Present-day practice in many electrochemical plants is to use constant current regulation. Constant voltage control is preferred in some applications. Combinations of both constant current and constant voltage can be applied and may be useful in start-up. Any method of regulation is likely to be expensive and should be applied only if natural regulation together with no-load taps are not sufficient. No-load taps can be applied economically over wide ranges, esspecially with the aid of Y-delta transformations and autotransformers.

For continuous control, four methods of regulation are available.

1. Step voltage regulators.
2. Induction regulators.
3. Self-saturating reactors.
4. Variable reactors.

Step Regulators

For relatively narrow ranges of control or wider ranges, when the coarseness of 32 step regulators is not objectionable, the standard step voltage regulator with modified controls to suit the particular application does an excellent job, with little effect on over-all efficiency and power factor. It is not applicable to jobs requiring fast response. It can be obtained in very large sizes and can therefore be used to control a large group of rectifiers feeding a single load. Used in this manner, it is naturally more economical than individual regulators for each rectifier and reduces the problems of control.

Induction Regulators

By their nature, induction regulators provide stepless control over wide ranges and affect efficiency and power factor only slightly more than step regulators. Because of the absence of contact mechanisms, their speed of response can be made faster than that of step regulators, but the ratings available are much lower and this hampers their application in large installations.

Self-Saturating Reactors

These can be applied in a manner similar to the commutating reactor of the mechanical rectifier. Actually, the entire unit becomes a self-saturating magnetic amplifier. They provide excellant stepless control and can be made extremely fast. The losses incurred in these reactors will generally be somewhat greater than in step or induction regulators of the same range. The power factor of the system decreases directly as the direct voltage is reduced. Small reactors can be used to good advantage to provide fine control between positions of the step regulator. Larger reactors can be used to obtain fast control over a wider range, in combination with a step or induction regulator to follow up and regain good power factor.

Variable Reactors

For small rectifier units in which efficiency and power factor are not important, the saturable variable reactor may be applied. These reactors increase the crest inverse voltage across the cells and must be adequately provided for in the over-all plan. The variable reactor does not greatly change the open-circuit voltage of the rectifier, but merely changes the slope of the volt-ampere characteristic. For some applications this may be an important advantage. In general, variable reactors cannot be recommended for large units, in that power factor and efficiency both are adversely affected, and the circuit crest voltage is increased.

Conclusions

Many factors must be considered in designing a rectifier installation. The interaction of various features on one another and the many different combinations that can be employed to meet specific requirements all need careful analysis. The optimum solution for a given case will not necessarily be the same for different manufacturers, even though the requirements are identical, because each manufacturer may have available to him different equipments and systems. Yet, there is a strong tendency for users of rectifiers to lay down a rigid specification to which all manufacturers must bid. In the author's opinion it would be to the user's distinct advantage to issue a functional specification, clearly stating the objectives in broad terms. Specifications should be detailed and rigid only in those particulars which are dictated by compelling circumstances. This will leave to the individual manufacturer the opportunity and responsibility of devising the best over-all plan compatible with his particular equipment design. By this means, the prospective user avails himself of the knowledge, experience, and skill of all the interested manufacturers. The final selection, based on an evaluation of the cost and characteristics of each proposed plan, has a vastly improved chance of being technically and economically correct.

In the case of the 3,000-kw rectifier previously mentioned, the author's company was fortunate in being able to work to a broad specification. Its success can be fairly attributed to the freedom of choice enjoyed. Of the many features incorporated, the following are the most important:

1. Forced inverse voltage division.
2. Forced current balancing.
3. Ultrahigh speed protection.
4. Liquid cooling in a truly closed system.
5. Use of deionized water for the coolant.

Forced voltage and current balancing made the coding of cells entirely unnecessary. No attempt was made to group cells in any particular pattern and in assembly they were picked entirely at random from the lot as received.

References

1. A New Voltage-Divider Circuit for Semiconductor Rectifiers, I. E. Dortort. *AIEE Transactions*, vol. 76, pt. II, July 1957, pp. 356–58.
2. Current Balancing Reactors for Semiconductor Rectifiers, I. E. Dortort. *Ibid.*, vol. 77, pt. I, Sept. 1958, pp. 452–56.

Discussion

E. J. Diebold (International Rectifier Corporation, El Segundo, Calif.): The present fluid state of development in large semiconductor rectifier units is hampered by a lack of definition. We should be grateful to the author for his lucid presentation which attempts to organize the great number of interconnected concepts.

Semiconductor rectifier units described by the author in this paper, and in the references, are outstanding in their small heat production compared with the power output of the unit. In this, the semiconductor rectifier element represents a great progress over the formerly used means of static rectification. It is further remarkable that these semiconductor rectifier elements may operate at very high temperatures, and that their operating temperature may vary within a very wide range without affecting the operation. Industrial power convertors, which are presently used, are noted for the need to keep the temperature of the gas discharge tubes at a low value, and that this temperature must be held constant. It is, for example, not possible to cool a gas discharge tube with a refrigerant, just to make sure that the elements do not overheat in case of overload.

The unique properties of having a small amount of heat to be removed and the high operating temperature of the rectifying element is an invitation to seek novel ways of cooling. The cooling of industrial devices, of course, is an old art. This does not mean, however, that it is well understood. I think, rather, that the art of cooling the new semiconductor devices lags behind the art of making them. The technique described in this paper is certainly extremely effective. There is no doubt that this method will provide excellent cooling under all circumstances. The question which I raise is, if this method of cooling is the most efficient in terms of expenditure in material and labor to manufacture the rectifier unit.

The mentioned drawbacks of the high-velocity air blast are true, if a high-velocity air blast is needed. If, however, an adequate heat exchanger is used, then air cooling does not need a high-velocity air blast. Because of the small amount of heat which is removed, the amount of air itself is not excessive and if low-velocity air is used the pressure drop is low. Thus, it is not necessary to use a large, high-power blower and

the major argument against air cooling is not valid.

The fact that air should be recirculated or cleaned before cooling is true in any electrical application. A well-integrated air-cooling system, which is taking advantage of the possibilities of insulation in air, can be just as safe and efficient as the method described by the author but may have advantages in first cost, floor space, and peace of mind, which warrant a more thorough investigation. It is particularly noteworthy that single-way rectification for mercury-arc rectifiers provides for water cooling at one electric potential, whereas semiconductor devices must be cooled at many different potentials which renders water cooling less safe.

If the need of water cooling is reconsidered, the necessity of using paralleling reactors at all the potential levels is eliminated. In this view our final conclusions are different from those reached by the author.

J. Chamulak (Pennsalt Chemicals Corporation, Wyandotte, Mich.): This informative paper on design features of large semiconductor rectifiers, to be considered by a manufacturer, has precipitated some thoughts of problems that a user of this type of equipment may have to contend with. These are as follows:

1. How would the number of service interruptions and down time for replacement due to failure of voltage-dividing components compare with a voltage-divider resistor installation and one using a voltage-divider transformer?

2. Recent published literature states that capacitors were installed in a rectifier to improve commutating duty. Is this feature to be considered in design?

3. Would parallel operation of semiconductor with mechanical rectifiers and rotary converters on only the a-c side require extreme care in selection of protective equipment?

Robert D. Lynch (Westinghouse Electric Corporation, Youngwood, Pa.): Mr. Dortort has covered rather completely the design problems which must be considered in the application of semiconductor rectifiers in large power installations. In the 125- to 500-volt range of applications, the silicon semiconductor has shown decided advantages over conventional rectifiers. At the lower voltages, germanium semiconductors with their lower losses have been applied quite successfully. However, contrary to early impressions, and as illustrated by Mr. Dortort's paper, semiconductors cannot be applied without considerable attention to the inherent thermal limitations and problems of voltage and current division.

The circuits which Mr. Dortort has developed to aid in his successful application of semiconductors indicate a rather complete solution to the various problems. It must not, however, be construed that there is only one solution to each of the problems involved. Not only will each customer have his own particular design requirements, but each equipment manufacturer will resort to techniques which are best suited to his past experience and present facilities. Only after several years of experience will the best techniques emerge and dominate the application of semiconductors. Germanium and silicon rectifiers have been applied quite successfully in a variety of ways. Cells have been placed in series with no forced division of voltage. Cells have been paralleled with no forced balancing of current, and various techniques are currently being used to properly protect against overvoltage and overcurrent surges.

There is no lack of knowledge concerning factors causing voltage failures of semiconductors. Exceeding the avalanche breakdown voltage of a semiconductor rectifying cell may result in a thermal failure from the excessive reverse current.

In addition, slight imperfections in a junction will cause a concentration of the excessive reverse current at a few points. Successive overvoltages may cause a gradual deterioration of the junction such that failure will not necessarily occur at the instant of the overvoltage surge. Voltage failures may also result from breakdowns across the surface of the junction. Impurities remaining from the manufacturing process, or slight imperfections of the surface are the cause of this type of breakdown. The mystery which generally exists results from the inability to determine the cause of failure except in laboratory-controlled tests. A failure most often results in an excessive current which destroys any evidence of the cause of failure. The low-energy surges which constitute the majority of high-voltage transients can be reduced quite successfully by a variety of techniques such as capacitors or nonlinear resistors. Installations experiencing high failure rates have been studied thoroughly and where overvoltage transients were detected, the addition of capacitors has completely eliminated further failures. It is not always possible to predict either the source or magnitude of overvoltage transients and the protection may have to be applied on a trial basis until more experience is gained.

Manufacturers of semiconductor devices have shifted back and forth in their opinions as to whether voltage dividers are required if cells are connected in series. Recovery times and the introduction of stray capacitance effects will cause unequal distribution of voltages during transient operations. Where several cells are connected in series it is now generally agreed that a form of voltage division is necessary.

Mr. Dortort's solution to the current overload problem has several advantages over conventional solutions such as current-limiting reactors. As he has mentioned, current-limiting fuses have been developed and used with success in many applications. However, a prejudice against fuses exists in some areas, and bolted faults on the transformer, as might be necessary with the short-circuiting switch, will require added mechanical strength in conventional transformers. Current-limiting reactors eliminate both of these problems but create other problems such as poor regulation and increased costs. The final decision as to what method is most desirable remains with the customer.

The major advantage of the silicon rectifying cell is its ability to operate at higher temperatures. Reliable operation has been obtained with convection and forced air cooling at these higher temperatures. Blower motors will decrease the over-all efficiency by less than 0.1%. Comparable losses would exist in circulating pumps. The liquid cooling of silicon rectifiers has not been generally adopted, but will certainly have its advantages in some applications. Again the customer must be the final judge as to which is more desirable.

Although Mr. Dortort has indicated a strong bias toward various design techniques, he has, in his conclusions, approached the general problem confusing the many potential users of rectifier equipment. The user is biased by his experience and is often reluctant to invest large amounts of money and endanger his production on equipment in the experimental stages. Each manufacturer has his own ideas as to which methods are correct, and these methods must be examined carefully and related to costs and particular needs. As previously mentioned, the best techniques will not be evident for several years, but there is no question that semiconductor rectifiers have emerged from the experimental stages. The installation of over 50,000 kw of silicon rectifiers in all types of applications in the last year and the successful operation of several units in electrochemical applications is evidence of this fact.

A. Schmidt, Jr., G. B. Farnsworth (General Electric Company, Lynchburg, Va.) and **J. P. Smith** (General Electric Company, Detroit, Mich.): The author has written an excellent paper expressing his views on this important subject. Because of the relative newness of their application to large rectifier systems, semiconductors provide several alternative design and application methods, most of which have merit. However, they are sometimes peculiarly suited to special or narrow applications and industries.

Large semiconductor rectifiers are applicable not only to the electrochemical industry, but to transportation and general industry. In these applications where the probability of external faults is relatively great the concepts of service continuity, fault protection, and device loading applying to electrochemical applications must be drastically revised. It is suggested that the title of the paper be modified to indicate that it applies chiefly to the chemical industry. Even in this industry the question of providing sufficient capacity so that 50% of the equipment may be shut down without curtailing production is a very special case and not usually economically feasible.

We have no fault to find with the author's current-balancing system, although his assumption that derating of coded cells is mandatory is erroneous. The physical problems of current-balancing reactors increase with the number of paralleled devices, but with increasing device ratings this should not be too serious a problem. A caution on the design of such reactors might be in order. Previous technical papers describing such a device have indicated that the reactors can be so designed to compensate for the maximum voltage spread at rated current on the assumption that the characteristic curve of these devices is essentially parallel above the rated point. On the contrary the typical volt-ampere characteris-

tics of two unlike semiconductor rectifiers show an increasingly wide spread as they go from a low to a high current value. Thus, for a properly designed reactor one must design for the volt-seconds necessary to parallel at the overload point, which may in some cases make them uneconomically large.

The practice of connecting a group of parallel-connected devices in series with similar groups rather than paralleling complete series circuits is open to severe criticism. When two series groups of parallel-connected devices are used the failure of any device in one group subjects all the devices in the other group to double voltage. It is one thing if the failure occurs in a faulty device under normal conditions and the stresses in the normal devices are within their safety factors. It is quite another thing if the failure occurs during a voltage transient or under conditions where the normal safety factor does not apply (this is a more likely case). Breakdown can take place in a matter of microseconds and no practicable protective device is likely to be devised which will adequately protect the normal cells in this situation.

The rectifier cell failure detection scheme described should be effective, but it requires individual fusing of each cell in a series string if failure of a single cell is to be determined.

The method of cooling best suited for semiconductors cannot always be justified on the basis of what is best and most efficient for the rectifier, but also what is needed by the application and industry in which the rectifier is used. Liquid cooling does provide more efficient heat transfer and often smaller size. However, the author's concern with corrosion, while justified when considering some unsatisfactory history, is not justified in light of the successful experience of the many properly designed water-cooled systems in operation in industry today.

Certainly the use of deionized water for a coolant is excellent, but it is questionable that one must go to this extreme to provide satisfactory service even over a 20-year period.

Requirements of different applications differ greatly and we agree with Mr. Dortort that the design factors should be carefully considered in any selection of a rectifier equipment. Any special circuit features should be evaluated realistically in the light of the degree of protection they provide at the expense of circuit complications. It is important also that the purchaser assure himself of adequate design margins in selecting semiconductor rectifier equipment.

J. W. Tracht (Pennsalt Chemicals Corporation, Philadelphia, Pa.): The author has pointed out the many factors which must be taken into account in the design and application of a rectifier installation. While the equipment manufacturer must give complete consideration to the details of construction, the user is primarily concerned with the reliability, conversion efficiency, and required capital investment of the rectifier installation. The relative importance of these three factors will vary between applications.

In most electrochemical plants it is mandatory that the conversion equipment be located adjacent to the process it serves and these process areas are usually subject to corrosive atmospheres. Forced-circulation air-cooling systems located in conditioned rooms have been utilized with reasonable success. However, unavoidable leakage of corrosive air into these systems, in view of the high volumes of air normally circulated, can result in severe corrosion over an extended period of time. Exposed contacts on instruments, control and metering equipment are especially vulnerable. Conductive deposits forming on high-voltage insulators have also been troublesome. Liquid-cooling systems are therefore normally preferred in areas where corrosive atmospheres prevail, and can also provide lower temperatures of operation and reduce the required installation space. It is important that individual rectifier elements be replaceable without the necessity of opening the liquid system. Supply and return piping of the cooling system should be so located and housed that accidental leaks will not reach electric equipment. Nonflammable liquid is mandatory.

Sufficient capacity should be built into the rectifier so that a reasonable percentage of the elements can fail without requiring an unscheduled outage of the complete unit. An alarm system to detect these failures is highly desirable.

During the development of new forms of conversion equipment it is often necessary to protect against weaknesses in basic elements of the rectifier. This usually involves the addition of equipment which complicates the installation and maintenance work and can contribute to a decrease in the reliability of the equipment. While this procedure is frequently unavoidable, the manufacturer of equipment should hold these additions to a minimum and design the equipment in such a way that, upon the availability of improved elements, they can be substituted and the auxiliary equipment removed. This is especially important where the auxiliary equipment inherently contributes to the losses and thereby reduces the over-all conversion efficiency of the unit. In the case of semiconductor rectifiers emphasis on the manufacture of cells with more uniform characteristics should therefore be stressed.

Likewise, during the development of new forms of conversion equipment the manufacturer should keep uppermost in mind his responsibility to the user for adapting improvements to the equipment which was originally supplied over a reasonable period of time.

Protection against external faults and overloads is of serious importance in semiconductor installations. High-speed operation is essential and from the users standpoint protection by open-circuiting is preferred to short-circuiting, since basically the application of a short circuit to a system, however short in duration or reduced in magnitude, inherently can affect other electric equipment on the plant system.

The desirability of close co-operation between the manufacturer and user during the early stages of design and application of equipment cannot be overemphasized. It is preferable that rigid design specifications not be initially laid down by either party. In electrochemical installations today the user can usually select from various arrangements and types of process equipment to obtain the production desired. This range of selection can involve a wide variation in the design of rectifier installations. The rectifier manufacturer has a responsibility to the user to acquaint him with the variations available, giving primary emphasis to the economic limitations and breakpoints between the various designs. With this information in hand, together with that on process equipment available, the user is then in good position to arrive at an over-all optimum installation. It is entirely possible that this over-all optimum installation may involve a design of rectifier installation which is not optimum in the eyes of the rectifier manufacturer. The rectifier installation will generally represent only about 10% of the total investment in the over-all electrochemical process installation.

H. Winograd (Allis-Chalmers Manufacturing Company, Milwaukee, Wis.): Mr. Dortort's paper is timely in view of the rapid increase in the application of semiconductor power rectifiers. Because of their higher efficiency, lower cost, and simplicity, semiconductor rectifiers are reaching into low-voltage application fields, below 200 volts d-c, for which mercury-arc rectifiers were not economical. They are also making rapid inroads into other application fields, which in the past were practically monopolized by mercury-arc rectifiers.

The new application problems, which were ably covered in the paper, result from the following limitations in semiconductor rectifiers as compared with mercury-arc rectifiers:

1. Their thermal capacity is small.
2. There is a sharp limit on the inverse voltage they can withstand.
3. The present limited range of current and voltage ratings per cell, which makes it necessary to connect cells in series and parallel.
4. There is no means for controlling the conduction of individual cells as is done with grid or ignitor control on mercury-arc rectifiers.
5. Unlike an arc back in a mercury-arc rectifier, which can be restored to service immediately after an arc back, a similar failure in a semiconductor cell is permanent.

Some of these limitations will undoubtedly be overcome by further development. The limitations of thermal capacity, maximum inverse voltage, and permanency of a failure seem to be inherent characteristics of semiconductor cells and are likely to remain.

Many experienced rectifier engineers, like Mr. Dortort, have been working on these problems, and different solutions have been proposed. There is considerable merit in a number of the ideas proposed in the paper. I have some doubts about several others.

With the series connection of two banks of parallel-connected cells, as described in the paper, a failure of a cell in one bank would impose momentarily double voltage on the cells of the other bank, which would endanger them.

Much can be said on the relative merits of water- and forced-air-cooling systems. For the 6-phase double-way rectifier circuit,

particularly when cells are connected in series, a water-cooling circuit would include a number of sections at different potentials, with insulating connections between them. This adds complications and may require more maintenance than an air-cooling system. We have manufactured semiconductor rectifiers with a recirculating-air-cooling system and an air-to-water heat exchanger for cooling the air. For silicon rectifiers, which can operate at a relatively high temperature, the objections mentioned in the paper to this type of cooling system are not serious. Very likely both water- and air-cooling systems will be used, depending on local conditions and the purchaser's preference.

I cannot agree with the doubts expressed in the paper regarding the benefits of surge-voltage protection for semiconductor cells. Since the cells are sensitive to overvoltage, suppression of voltage surges by arresters, such as capacitors properly applied, should be beneficial.

One important factor in the application of semiconductor rectifiers, which was not mentioned in the paper, is the effect of harmonic components in the a-c supply voltage on the d-c output voltage of individual rectifiers.[1] When two or more parallel 6-phase rectifier sections or units are displaced in phase to produce combinations of 12 phases or more, as is usually done in electrochemical applications, harmonics in the alternating voltage could cause current unbalances among the parallel rectifiers. Unbalances of 10% ro 15% are encountered frequently. This is sometimes aggravated by the presence of capacitor banks on the a-c system, which could generate harmonic voltages through resonance with the reactance of the line or transformers. In mercury-arc rectifier installations, the unbalances can be corrected by phase control, using grid or ignitor control. In semiconductor rectifier installations such unbalances would be more serious, because of their limited load capacity, and would have to be corrected by means of adjustable reactors or fine transformer taps for the individual 6-phase rectifiers. In some rectifier installations the unbalance shifts as a result of a-c system changes or switching of capacitor banks, which change the magnitude or phase position of the harmonics in the alternating voltage. Automatic current-balancing means would have to be used in such cases.

Operating experience with semiconductor rectifiers has been limited so far, especially at 250 volts d-c or higher. A number of different solutions will undoubtedly be found for the application problems encountered, and it will take some time before optimum designs, proved by experience, are attained. Papers of this type are useful in pointing out the technical problems involved in the application of the new rectifiers and in generating discussions.

Keen competition among manufacturers is likely to produce improvements and cost reductions. There is a risk that such competition will result in some unwise short cuts and reduction of safety factors beyond the safe limit. Purchasers' engineers should be on guard against such practices. The technical organizations could also provide a measure of protection by setting up sound quality standards. Such work is in progress.

Reference

1. Harmonics and Load Balance of Multiphase Rectifiers, R. D. Evans. *AIEE Transactions*, vol. 62, 1943, pp. 182–87; *discussion by* H. Winograd, pp. 400–01.

I. K. Dortort: The many searching discussions presented on this paper are a true measure of the interest in the subject. I am very grateful. Let me hasten to agree that there are many possible solutions to the problem involved and that time will eventually sift out the less satisfactory. This does not imply that any one solution will ever be right for all cases. We are hopeful that the basic precepts will stand the test of time. Mr. Tracht's views strengthen this hope.

There appears to be complete unanimity on the need for adequate cooling, the desirability of effective current and voltage division, and the importance of protection. Beyond this area of agreement the field is "wide open." I will try to separate the issues and clarify our position with respect to each, rather than to attempt a detailed reply to each discusser.

Series Connection of Paralleled Blocks of Cells

Regardless of the cooling means employed, the mounting of many cells on paralleling busses provides the immediate advantages of a great reduction of the number of mounting details, copper, reactance, and space requirements. Even with no provisions for inverse voltage distribution, the statistical distribution of reverse currents among many paralleled cells will assure a fair distribution of inverse voltage between blocks. Its one possible disadvantage is the loss of the statistical averaging of forward drops obtained in parallel series strings. This factor is almost negligible when there are only two or three cells in series. Current distribution in a block without reactors is no worse than in rectifiers with no series-connected cells. The addition of balancing reactors is not required by the series connection of blocks. The physical arrangement merely makes it easier to add reactors if desired.

It has been stated that derating of cells paralleled without forced current division is not mandatory. That is certainly true. If sufficient care and precautions are taken to assure accurate current division by means other than reactors, no derating is required. It remains to be proved that this can be accomplished economically, both in the initial installation and subsequent cell replacements. Whether we speak of derating per se, or conceal the derating factor in the margin of safety does not change the cell loading or the number required.

The reduction in the number of reactors for paralleled series strings, as suggested by Mr. Diebold, is a reduction in numbers only, not in volt-seconds, unless there are many cells in series. With the particular structure we are using, the total size and cost would probably be somewhat reduced. This gain would not outweigh the disadvantages. Transformers or resistors can be employed with either the series-block or parallel-string system. As a matter of fact, since the potential-divider transformers under normal operating conditions supply only the unbalanced leakage currents of the cells, it is quite feasible to connect a single transformer winding to a large number of cell junction points in individual series strings through relatively high resistances, without appreciably upsetting the voltage balance, yet keeping each string effectively isolated from the others.

In reply to Mr. Chamulak's specific question on service interruptions due to failure of either voltage-divider transformers or resistors, this is a matter of the quality of the individual components used and the number of components required, plus the number and type of connections. On the basis of these considerations, and the susceptibility of resistors to the corrosive atmospheres around chemical plants, it is reasonable to expect that transformers used with series blocks will give less trouble than resistors used with parallel strings of cells.

A fuse in series with a string of cells does not protect these cells in case one of them fails, it cannot blow until all of them have failed. In the series-block arrangement, it is possible to fuse each cell individually when used with a particular design of voltage-divider transformer. It has been implied that this is an economic disadvantage. It must be obvious, however, that the first cost of these additional fuses is lower than special cell failure-detecting devices applied to each individual cell, and far less complicated. In the matter of replacement cost, the individual fuses are less expensive because of their lower voltage rating and a great deal less expensive than the number of cells that must fail in a series string to blow a single fuse.

The objection has been raised that when blocks of paralleled cells are connected in series, the failure of any one cell will subject all the cells in the other blocks to an overvoltage; whereas, if the cells are connected in individual series strings, the overvoltage is applied only to the few cells in series with the one that failed. On its face, this appears to be a valid objection. Let us consider first only those failures which do not occur during voltage surges.

It is true that the total circuit voltage will be impressed across the remaining good blocks of cells and this overvoltage will be maintained until the protective devices have functioned. If the protective device is a high-speed bypass switch (burdened with the name short-circuiter), the overvoltage will be maintained for approximately 100 microseconds. If individual cell fuses are used, the overvoltage will last perhaps 8 to 10 milliseconds; but even during these short intervals, the voltage across the remaining cells, even with only two blocks in series, will not exceed the PIV rating of the cells which in present practice is made two or three times as great as the peak circuit voltage.

If individual cell fuses are used, the failed cell is removed from the circuit and the current divided between the remaining cells of the block. The current distribution in the other blocks is not affected. Now, let us consider what happens in a system of paralleled series strings. The failure of one cell increases the voltage across the other cells in the string and this increased voltage is maintained until all the cells in the string have failed, or until remedial action has been taken. Assuming no remedial action

is taken and the fuse blows, the current in all other cells on that phase will be increased. It seems then that at least for failures under steady-state conditions, all the cells in the string in which failure has occurred are jeopardized for an indefinite period of time by the increased inverse voltage until they have failed and a fuse blown, and all the other cells in that phase are then jeopardized by increased current until remedial action is taken. In contrast, the series-block arrangement with individual fuses provides immediate voltage protection for all the cells in the block in which failure occurs, producing overvoltage in the other blocks for only a fraction of a cycle. Overcurrent is confined to the one block in which failure has occurred.

Transient Overvoltage Protection

Now to analyze the advantages or disadvantages of the two systems with respect to transient voltage surges. In the series-block arrangement, if failure of a cell occurs during a voltage surge, there is no question that the full surge will then appear across the remaining series blocks. If, however, these surges cause thermal failure through the mechanism of avalanche, or, if failure is caused at some later time due to cumulative deterioration caused by repeated surges, as suggested by Dr. Lynch, the effect on other blocks will be as described in the preceding section. The number of cells jeopardized, and the degree, will not be so great as in the case of the series-string arrangement.

Dielectric breakdown due to impurities or imperfections across the edges of the crystal will undoubtedly occur before the surge has passed and the condition of overvoltage on series blocks suggested by Mr. Schmidt, Mr. Farnsworth, and Mr. Smith will prevail. By the same token, the possibility of unequal voltage distribution, especially during transient conditions, will under the same circumstances endanger the weaker cells in all paralleled series strings. The degree of jeopardy is further increased by the difficulty of obtaining proper voltage division during transient conditions as pointed out by Dr. Lynch. Any resulting additional failures increase the hazards of overcurrent on the remaining cells. On the other hand, all the good cells in a block are protected by the first cell to fail. The cells in the seriesed blocks will similarly be protected by the first cell that might fail in them.

It is encouraging to be informed that the causes of overvoltage failure are all known and understood. It is easy to see how progressive deterioration, the inability to predict, observe, or measure surges that do occur, and the destruction of evidence when cell failure occurs, have mystified many of us and produced a feeling of uncertainty regarding protective measures. It did not mean to express any doubt that capacitors and nonlinear resistance elements are beneficial. We use them. I would like to believe that they are a sure cure, but cannot bring myself to this point when told that these protective elements may have to be applied on a trial basis until more experience is gained.

Experience to date does not indicate any need for serious concern about the propagation of overvoltage failures in the series-block system. Neither the individual fuses, nor the high-speed protective switch would be fast enough to avoid the appearance of full surge voltage across the remaining series-connected blocks, after a cell had failed in any one of them during a surge. Both protective systems have been used, backed up by capacitors and nonlinear resistors. No simultaneous or sympathetic failures have occurred.

The question will naturally be asked why we do not use capacitors large enough to drown all surges and forget this problem. The solution is not that simple. It is quite possible to set up resonant oscillations between the capacitors and the leakage field of the rectifier transformer with resulting inverse peak voltages across the cells as much as double the peak of the fundamental wave.

When capacitors used for surge-voltage protection approach or exceed a value which in resonance with the commutating reactance produces a half-cycle period of the order of magnitude of the overlap angle, their effect on regulation and primary current waveshape must be taken into account. Small capacitors sufficient only to suppress low-energy surges are not likely to exert any influence on other design factors.

High-Speed Protectors

Disfavor has been expressed in several discussions of our use of a short-circuiter to protect the cells in the event of faults, rather than reliance on some form of high-speed breaker. Had we christened this device a high-speed protector it might not have met such opposition. As stated in the paper, there is no need to impose a dead short circuit on the rectifier transformer to protect the cells. We insert enough impedance in the connections to the protective device to permit only enough current to pull down the transformer voltage to a value which will not deliver more current to the fault than the cells can handle during the time it takes the breakers to clear. The stress on the transformer is very moderate and well within the capability of standard power transformer construction. The necessary impedance in the protector circuit is very easily obtained by using properly dimensioned iron straps instead of copper bussing.

Cooling

The subject of cooling seems to be the most controversial. While we are convinced that water cooling is best for most applications, we will shortly ship a large rectifier for a very special application, using silicon cells cooled by forced air. Engineering studies on this project have convinced us that:

1. Considerable margin of safety is sacrificed in using air unless cell loading is reduced.
2. High-velocity air is required for an acceptable margin of safety unless extended cooling surfaces are increased, making the equipment larger.
3. Power required for blowers is many times greater than that for water pumps.
4. Considerably more building space is required for air-cooled units.
5. Maintenance will not be lower for air-cooled than water-cooled units.

Even though we have taken advantage of the compactness of the series-block arrangement, the rectifier compartment is 25% larger than would be required for a comparable water-cooled rectifier. This does not take into account space requirements for blowers, air-to-water heat exchangers, and ducts, which would be required in other air-cooled installations.

We find that on the basis of 35 C (degrees centigrade) air for the air-cooled unit, or 35 C water for the water-cooled unit, the air-cooled unit in this case would require auxiliary power of at least 0.6% of the rating of the rectifier, if the air had to be circulated through heat exchangers and ducts. The water-cooled unit, on the other hand, would require only 0.016%. Under these conditions the junction temperatures are estimated at 122 C for the air-cooled unit and 96 C for the water-cooled unit. The temperature margin of safety on the water-cooled unit would be 38% greater than for the air-cooled unit. At higher cell loadings the difference would be even more pronounced. Actually, of course, if the air had to be cooled in air-to-water heat exchangers, the inlet temperature to the rectifier would be considerably higher than 35 C and the difference in junction temperatures would be still more pronounced. More likely the water temperature in this location would be less than 35 C, but the difference would be about the same. A reduction in blower horsepower could be achieved by operating the cells at higher junction temperatures and lower margins.

Obviously, deionization of water is not the only means of obtaining a satisfactory cooling liquid. As Mr. Schmidt has properly pointed out, there are many mercury-arc rectifiers in service today using sodium dichromate in their cooling water, and electrolytic targets to alleviate electrolysis at hose nipples and fittings. The installation and maintenance of targets in multichannel, multipotential systems is onorous. In the case of a double-Y-connected rectifier with a single cell in each leg, we could use the system that has been in vogue so many years. It certainly would not be a desirable system for bridge-connected rectifiers with two or more cells in series. Considering the low cost, small space requirements, and simple maintenance of the small deionizer required in most cases, there seems to be no good reason for not using a deionizer whenever water cooling is employed.

In comparing the maintenance of air- and water-cooling systems, it is well to keep in mind the point made by Mr. Tracht that when handling large volumes of air in and around chemical plants, unavoidable leakage of corrosive gases into the closed air system and into the rectifier room will occur. These gases will attack control and metering equipment and may produce conductive deposits on insulation. Cleaning of cooling surfaces, insulation, heat exchangers, etc., must be considered in the over-all maintenance program. Water-to-water heat exchangers, pumps, and deionizers available today in the ratings required for a complete water-cooled rectifier unit, or a section of a rectifier, present minor maintenance problems and are small enough to permit installation of 100% spares, if required, without any great additional cost.

Mr. Diebold makes the point that the series-block arrangement is not needed when

air cooling is used. I have previously mentioned that cooling is by no means the only reason we use this arrangement and it is in fact being used on the large air-cooled rectifier now in process, after a careful analysis of all factors including cost and space requirements. We cannot agree that the air-cooling system is more efficient, lower in cost, and requires less floor space, when applied to large rectifiers.

Application

I admit the title of the paper is not fully descriptive in that it was written with the chemical industry in mind. Nevertheless, I hope it will be found useful in other applications. The problem of external faults has been successfully met in one chemical installation and there is no technical reason why it cannot be done in other applications. There is more to the problem. It will be of little benefit to attempt to control severe load peaks and short circuits by prohibiting them on the name plate of the rectifier. Cell loading must be matched to cell capabilities and high-speed protection designed to protect the cells under all conditions. Traction and industrial rectifiers will undoubtedly cost considerably more than chemical application rectifiers of the same nominal rating.

Perhaps one of the most difficult evaluations that must be made in planning any new installation is that of stand-by capacity, connected or idle. This evaluation must consider the number of units, functional sectionalizing of units, and even the number of cells connected in parallel. With the cell ratings presently available, it is quite reasonable to connect enough cells in parallel to permit operation at full load with one, two, or even three cells disabled in any one bank. It does not make much sense however, if sufficient units or sections are available, to permit operation with less than the full number of cells, for any time longer than necessary to arrange a shutdown for replacement. It may be that provisions for cell replacement without shutdown will become standard practice. It remains to be seen, however, whether this is economically justifiable; especially in view of the low and decreasing incidence of cell failures.

The use of a functionally sectionalized rectifier in which one half of the unit is shut down in the event of cell failure might not always be justifiable if one section represents 50% of the total bus capacity. We do not advocate this procedure; but even the occasional loss of 50% of capacity for a few minutes, or even a few hours, has little effect on the yearly production of a continuous process chemical plant. If a short-time interruption can be tolerated, a significant reduction in first cost and maintenance costs can be achieved.

Mr. Tracht informs us that in many chemical plants the conversion equipment represents only 10% of the cost of the plant and therefore the design of this equipment may have to be less than optimum because of its subordination to the other 90%. I find this rather hard to understand. Where power is a major product cost, and continuity of service a necessity, the conversion equipment may assume importance far beyond its dollar value.

In reply to Mr. Chamulak's third question, paralleling of semiconductor rectifiers with mechanical rectifiers and rotary converters, or for that matter, mercury-arc rectifiers, on the a-c side only, poses no problems in protection unless these machines or their switchgear produce high-voltage transients in the a-c network. Paralleling on the d-c side requires careful analysis. It can and has been done.

Cell-Failure Alarm

Automation can be a wonderful thing. Today there is a trend to build cybernetics into rectifiers to keep track of cell failures and to make decisions under various combinations of circumstances: when to wake up the operator and when to shut down the machine. These systems can become complicated and expensive, and may prove an annoying source of maintenance. Some systems require a large number of small wires from the power circuits to control cabinets and they constitute a source of danger.

If this policy is wrong, it is up to the users to correct it. Should the manufacturers provide only indicating-type fuses to disconnect and locate bad cells? Should an alarm be added for cell failure? Or, is it necessary to give an alarm for failure in any branch of the rectifier, count the number of failures, evaluate the positions of the failures, and after proper decision making, trip the breakers? To me it seems the last step is going too far. Nothing has ever been made foolproof; but unless the purpose is sabotage, no one will deliberately wreck a rectifier.

What are the probabilities and extent of damage in the event of complete and total negligence? After the first fuse blowing and alarm, it may be months before the next, and the next, failure. It is conceivable that eventually all cells in a given phase will fail before any remedial action is taken. If the rectifier is supplying an isolated load, the direct voltage and load will drop, finally calling attention to this unpleasant condition. In multiunit installations the change may not be noticeable. Would the maintenance and operating organizations in any plant permit such a condition to come about? Is the remote possibility of such an occurrence worth the extra cost, complexity, and maintenance required to avoid it?

Harmonics

Mr. Winograd has done the industry a service in calling attention to the possibility of wandering unbalance caused by harmonics. In mercury-arc rectifiers, the condition can be corrected by a small amount of phase control. Simple and economical means of phase retard for semiconductor rectifiers have been developed and tested and provide a small range of accurate and fast control. They are available to take care of the condition mentioned, but will not be installed unless required. When installed, local conditions will decide if correction is to be manual or automatic.

Summary

Some of the problems discussed will decrease in importance as cells are improved. Many research and development organizations are working in this area, prodded by keen competition. Eventually this will result not only in better quality, but in the greater uniformity desired by equipment manufacturers and users. A further responsibility rests on the cell manufacturers to provide factual, useful information on the cells presently available and on the new ones appearing on the market. In self-interest and for the benefit of the entire industry, an effort should be made to make new cells interchangeable with the old. The equipment manufacturer has the responsibility of designing the most reliable units to use the quality of cells currently available. This should be done with the least complication possible, but again, the evaluation requires a careful analysis of all the interrelated design features discussed in the paper. It is impossible to foresee future developments, but in this fast-moving field, provision should be made for anticipated improvements, if practicable, and full cooperation given the user in modernizing existing units.

The industry as a whole must assume the responsibility of writing standards. Committees in AIEE, National Electrical Manufacturers Association, and International Electrotechnical Commission are working toward this end. We hope they will all try to achieve standards to ensure the user quality products, and even more, a full knowledge of what he is getting, and at the same time, permit as much latitude as possible in the development of a new field.

The particular design features and practices stressed in this paper are not the work of one man. They are the result of the efforts and interchange of ideas of the engineering staff of I-T-E's Transformer and Rectifier Division. Many of the points discussed here have been considered and evaluated in the process; thus, I extend my thanks and appreciation.

Analysis of Ignitron Rectifiers for Reversing-Mill Drives

C. G. HAGENSICK
MEMBER AIEE

E. J. CHAM
MEMBER AIEE

A REVERSING-MILL DRIVE for rolling-mill service for rolling blooms, slabs, or plates requires application of d-c motors operated under a special load cycle. The motor must be started, accelerated, operated under load, and decelerated many times as the metal which is being rolled is passed back and forth through the rolls. On alternate passes the motor is operated in the reverse direction. The time required for each pass is a few seconds. Typical motor ratings range from about 3,000 hp (horsepower) to about 12,000 hp. The required time for a speed change from base speed in the forward direction to base speed in the reverse direction is about 1 to 2 seconds. Several recent applications of this type have been built with a reversal time of 1 second.

The conventional method of supplying motors for reversing-mill drives is the Ward Leonard system, which utilizes a motor-generator set to supply variable direct voltage to the motor rated at 750 volts or less for domestic applications. Another method, analyzed in this paper, is to utilize electronic power converter equipment containing mercury-vapor tubes such as ignitrons; this method has been studied and applied in Europe using continuously excited electronic power converters.[1-4] With the motor-generator set, the direction of power to the drive motor is controlled by the field of the d-c machine. With an electronic power converter the direction of power is determined by the ignition angles of the tubes so as to operate the tubes either as a rectifier or as an inverter.

The purpose of this paper is to investigate the feasibility and possible use of ignitron converters for reversing-mill drives. This paper includes discussion of the principles, power circuits, control methods, protection, and tube considerations for this type of application.

Advantages and Disadvantages of Electronic Power Converter Drive

In applying electronic power converter equipment for the power supply in this type of application, the following well-known advantages of rectifier equipment over rotating equipment are realized:

1. The electronic power converter is a static device and requires no moving parts except the water pumps on all units and the vacuum pumps on pumped units. This reduces the maintenance work and increases availability.

2. The foundation and ventilation requirements are much less for electronic power converter equipment, thus appreciably reducing the installation time and expense.

3. The efficiency is higher for electronic converter equipment, particularly at light loads.

4. The control of voltages and response to changes can be made faster since there is no inherent inertia in electronic converter equipment.

The following are disadvantages of electronic power-converter equipment as applied to this type of application:

1. The load surges due to the variable loading and regeneration are reflected instantaneously on the power system, there being no flywheel or inertia effect as with rotating apparatus.

2. Because of the unidirectional current characteristics of electronic tubes, additional equipment is needed to accomplish power reversal without current reversal in a given electronic power converter. This additional equipment may include reverse-connected converter equipment, or it may include equipment to reverse rapidly the motor armature or field.

3. To use an electronic power converter to its best advantage it is necessary in most cases to use a new and less desirable type of main motor design. This motor may be designed for a voltage higher than the conventional maximum of 750 volts or it may be designed with a laminated frame, or both. The need for higher voltage is quite serious to the motor designer, and the use of a laminated frame may increase costs appreciably.

4. The power factor on the supply system is lower due to electronic converter operation with appreciable phase control.

Paper 58-218, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted May 27, 1957; made available for printing November 19, 1957.

C. G. HAGENSICK and E. J. CHAM are with Westinghouse Electric Corporation, East Pittsburgh, Pa.

Principles of Rectification and Inversion

Rectifier and Inverter Operation

Ignitron rectifiers have been used extensively in various electronic power converter applications. Most of these applications are for the conversion of power from an a-c circuit to a d-c circuit, or rectification. Some applications utilize electronic power converter equipment for the transfer of power from a d-c circuit to an a-c circuit; this is called inversion. In electronic power converters utilizing gaseous tubes, such as ignitrons, the basic method to control transfer of power is to control the start of the conduction of each tube. In ignitrons this is accomplished by passing a pulse of current through the ignitor, which then fires and initiates the cathode spot and subsequent pickup of the tube.

Fig. 1 shows the power circuit of a delta 6-phase double-Y rectifier, which is a commonly used circuit. The a-c winding of the rectifier transformer is connected by means of the a-c circuit breaker to the a-c supply. The d-c winding of the transformer consists of two 3-phase Y groups with each phase connected to the anode of an ignitron. The anode breaker shown is used on some but not all rectifier applications. Anode breakers are used for the larger ratings or for essential service applications. The interphase transformer, which is connected between the neutrals of the two Y groups, is used to withstand the instantaneous voltages between these groups so that each group may operate as a simple 3-phase rectifier. The rectifier load is connected between the mid-point of the interphase transformer and the cathode of the rectifier.

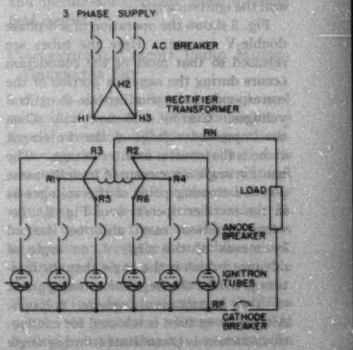


Fig. 1. Typical delta 6-phase double-Y rectifier circuit

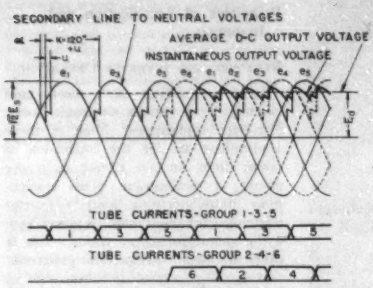


Fig. 2 (left). Operation of 6-phase double-Y rectifier

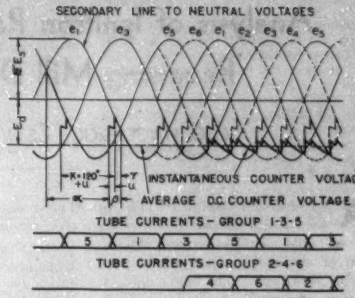

SECONDARY LINE TO NEUTRAL VOLTAGES
INSTANTANEOUS COUNTER VOLTAGE
AVERAGE DC COUNTER VOLTAGE
TUBE CURRENTS - GROUP 1-3-5
TUBE CURRENTS - GROUP 2-4-6

Fig. 3 (right). Operation of 6-phase double-Y inverter

The cathode breaker is provided to connect the output of the rectifier to the load. When anode breakers are not used, the cathode breaker is also used to clear arc backs. The circuit in Fig. 1 may be used for both power rectification and inversion; its actual operation at a given instant is dependent upon the ignition angles of the tubes with respect to the circuit voltages.

Fig. 2 shows the voltage and current wave forms of a 6-phase double-Y rectifier. Operation is shown for a small amount of ignition delay and for commutating angle u. The ignition angle α is measured from the positive crossing point of the instantaneous d-c winding line-to-neutral voltages. In rectifier operation most of the conduction period κ of a given tube is in the positive portion of the corresponding d-c winding voltage. The commutating angle u is the angle during which the current is transferred from one phase to the next phase in the same simple rectifier group; it is dependent on the current transferred, the circuit voltage, the commutating reactance, and the ignition angle α.

Fig. 3 shows the operation of a 6-phase double-Y inverter. Here the tubes are released so that most of the conduction occurs during the negative portion of the corresponding d-c winding line-to-neutral voltages. Current will flow only when the impressed voltage of the d-c circuit exceeds the inverter countervoltage. The ignition angle α is measured from the same positive crossing point of the voltages as in the rectifier operation of Fig. 2. Inverter ignition may also be defined by means of the angle β, or angle of advance, which is the angle that the ignition precedes the negative crossing point of the instantaneous phase voltages. After a given tube is released for conduction, current is commutated during angle u. The angle γ between the end of commutation and the negative crossing point is called the margin angle and includes the time required for tube deionization. If the margin angle γ is too small so that the tube deionization is incomplete at the negative crossing point of the voltages, the tube of the outgoing phase then has positive voltage impressed across it and may again conduct, which will result in an inverter fault.

Voltage Characteristics

The average value of the output voltage of a rectifier under load is indicated by E_d in Fig. 2. The average value of the d-c countervoltage of an inverter under load is indicated by E_d in Fig. 3. The voltage value of the d-c output or countervoltage of a power converter is

$$E_d = E_{d0} \cos \alpha - E_x - E_r - E_a \qquad (1)$$

where E_{d0} is the theoretical average no-load voltage, E_x is the average voltage drop due to commutating reactance, E_r is the average voltage drop due to resistance, and E_a is the average element drop. These quantities are determined from the following:

$$E_{d0} = \sqrt{2}\, E_s \frac{p}{\pi} \sin \frac{\pi}{p} \qquad (2)$$

$$E_x = \frac{p}{2\pi} I_c X_c = E_{d0} \frac{\cos \alpha - \cos (\alpha + u)}{2} \qquad (3)$$

$$E_r = \frac{P_r}{I_d} \qquad (4)$$

E_a = value determined from arc-drop measurement

For a delta 6-phase double-Y circuit, $p=3$ and $I_s = I_d/2$. Therefore,

$$E_d = 1.17 E_s \cos \alpha - 0.239\, I_d X_c - \frac{P_r}{I_d} - E_a \qquad (5)$$

An alternate expression for the inverter countervoltage in terms of the advance angle is

$$E_d = -\left(1.17 E_s \cos \beta + 0.239\, I_d X_c + \frac{P_r}{I_d} + E_a\right) \qquad (6)$$

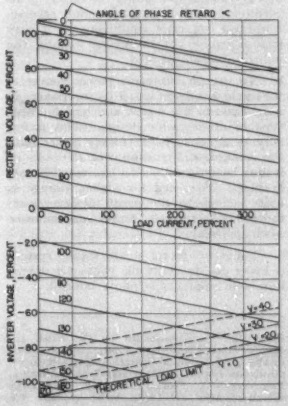


Fig. 4. Regulation curves of power converter, with 8% regulation

Fig. 4 shows typical voltage-regulation curves of an electronic power converter showing both rectifier operation and inverter operation for different ignition delay angles. The positive values of direct voltage indicate rectifier operation, while the negative values indicate inverter operation. These regulation characteristics consist of a family of parallel drooping curves with the highest curve for $\alpha = 0$ and the lowest curves for higher values of α. These curves do not show the light-load voltage rise which is characteristic of this circuit; however this is unimportant for analysis of this type of application. The values at zero current are proportional to the cosines of the respective delay angles α. The slope of the curves is determined mainly by transformer reactance. The inverter regula-

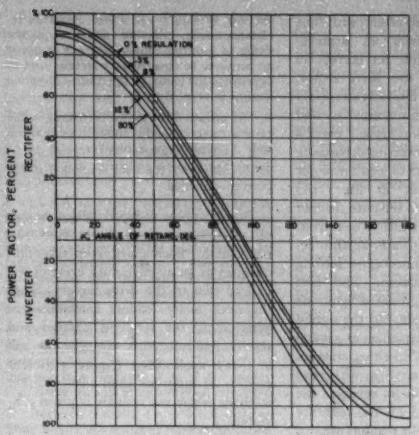


Fig. 5 (left). Power factor of 6-phase rectifier or inverter at rated current, neglecting exciting currents and losses

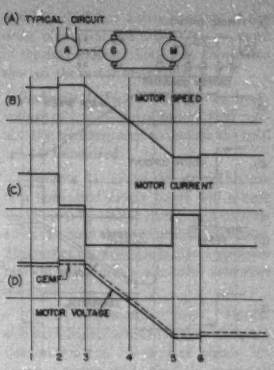


Fig. 6 (right). Circuit and operation of conventional reversing-mill drive using motor-generator set

tion curves terminate at the theoretical load-limit line at the bottom of Fig. 4. Along this line the commutating angles are exactly equal to the advance angles and there is no margin angle for deionization. In actual operation it is necessary to operate with sufficient margin angle to provide for tube deionization; a practical inverter load-limit line is then one with the desired margin angle γ such as those shown above the theoretical load-limit line and parallel to it.

Energy Flow and Power Factor

The direction of energy flow in an electronic power converter is determined by the polarity of the voltage, which in turn can be adjusted by control of the ignition angles. Current cannot be reversed in a given converter because of the blocking action of the electronic tubes; therefore, power reversal can be accomplished only by reversing the voltage. Hence the operation of the electronic power converter differs from that of a d-c machine, in which the current reverses to accomplish reversal of power, or regeneration. In applying electronic power converters to reversing-mill drives which require regeneration, it is necessary to make provision in the circuits so that the converter tubes will carry positive current and will have the circuit voltages impressed in the proper polarity upon the converter when it is desired to regenerate. The three basic methods of accomplishing regeneration and reversal with electronic power converters are described in the following section.

The use of phase control to obtain adjustment of voltage over the range of rectifier and inverter operation has a disadvantage of reduced power factor as compared with operation without phase control. Fig. 5 shows the power factor of any 6-phase power converter for various values of ignition delay angle α. These curves are for different values of direct voltage regulation, which is the ratio expressed in per cent of the voltage drop between no load and full load to the full-load voltage for rectifier operation without phase delay. The curves terminate at the lower right at points corresponding to the theoretical inverter load limits.

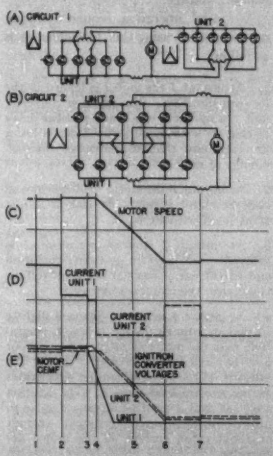


Fig. 7. Circuits and operation of drive with dual ignitron converters for motor armature

Power Circuits for Reversing Drives

General

This section discusses various arrangements which can be used to operate reversing-mill drives. Figs. 6 through 9 show typical circuits and characteristics to illustrate motor reversal with each method. The numbered instants at the bottom of each figure indicate steps in the respective sequences of operation. The intervals between these steps are not indicative of relative time values, since in some cases these intervals are exaggerated to assist in understanding the operation. These sequences of operation show motor reversal from base speed forward to base speed reverse without load on the motor during any part of the reversal. Although this type of operation is described, it is pointed out that actual mill cycles for any circuit also include acceleration and deceleration with load and operation above base speed by using field control.

Conventional Circuit Using Motor-Generator Set

Fig. 6 shows a typical circuit and method of operation of a conventional type of reversing-mill drive using a motor-generator set. The a-c machine A is connected to the a-c supply circuit, and the d-c machine G is connected to the armature M of the drive motor. In normal operation the a-c machine operates as a motor to drive the d-c machine, which in turn operates as a generator to provide the proper voltage on the drive motor. When it is desired to decelerate the drive motor, the generator voltage is reduced by field control so that it is lower than the counter-emf (electromotive force) of the drive motor. The direct current reverses

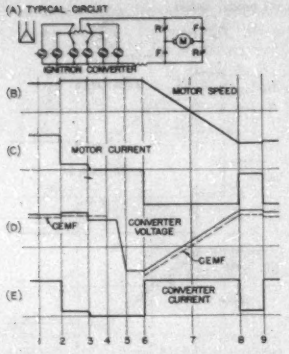


Fig. 8. Circuit and operation of drive with single ignitron converter and reversing contactor

and power is transferred from the drive motor to the d-c machine of the motor-generator set. This d-c machine now operates as a motor to drive the a-c machine as a generator, thus resulting in regeneration. Further adjustment in the field of the d-c machine can be used to bring the drive motor to rest and to start it in the reverse direction. A flywheel is often used with the motor-generator set so as to minimize power pulsations on the supply system. The a-c machine then may be a wound-rotor induction type with provision for speed control in the secondary circuit.

A typical operation sequence during the motor reversal for the circuit of Fig. 6 is as follows:

1. At instant *1* the motor *M* is carrying load in the forward direction.

2. At instant *2* the load is removed and the motor is now operating at a constant speed without load.

3. During the interval *3* to *4* the field is reduced on the d-c machine *G* of the motor-generator set. The direct current reverses and power is transferred from the drive motor *M* to the motor-generator set.

4. At instant *4* the drive-motor speed is zero and regeneration is complete.

5. During the interval from instants *4* to *5* the field is increased in the opposite direction on the d-c machine *G* of the motor-generator set. The drive motor *M* is accelerated in the reverse direction.

6. At instant *6* reverse load is applied with the drive motor at the desired speed.

Dual Ignitron Converters for Motor Armature

One method to provide for regeneration and reversal in a drive employing an electronic converter system is to use two ignitron converters with one reverse-connected. Such an arrangement provides a circuit for current in either direction for the drive motor. This method has been applied extensively in Europe using continuously excited tubes.[1,2] Fig. 7 shows typical circuits and operation of the reversing drive using dual ignitron converters to feed the motor armature. The circuit in Fig. 7(A) shows a method using a separate rectifier transformer for each ignitron converter. The circuit in Fig. 7(B) shows an arrangement with a single transformer feeding two groups of ignitrons. The first arrangement has two groups of ignitrons each at common cathode, but the second arrangement has the ignitrons at several cathode potentials. With either arrangement reactors are required to withstand the instantaneous voltage between the ignitron units and to limit the resulting circulating current. The reactors are also required to limit the ripple in the motor current, especially at the conditions of high rectifier delay and with inverter operation.

This electronic converter circuit must be completely symmetrical with respect to voltages and number of tubes for each direction of drive-motor rotation, as the same amount of rectification and inversion is required for each direction of rotation. No dissymmetry is possible such as in a railway substation application in which there can be fewer inverter units than rectifier units and in which the inverter units can be operated with a higher d-c winding voltage.

A typical operation sequence during motor reversal using this arrangement is as follows:

1. At instant *1* the motor is loaded in the forward direction. The ignitron converter unit *1* is operated at a small rectifier delay so that its average output voltage is lower in magnitude than the inverter counter-voltage of unit *2*.

2. At instant *2* the motor load is removed and the motor is now operating at a constant speed without load.

3. At instant *3* ignitron converter *1* is delayed to an inverter operating condition and its current is blocked when its voltage is less than the motor counter-emf.

4. At instant *4* unit *2* is advanced slightly so that its countervoltage is less in magnitude than the motor counter-emf. Current now is carried in converter *2* during the interval *4* to *6*. The motor decelerates during the interval *4* to *5*. The ignition angles of unit *2* must be advanced so as to provide approximately constant current at about rated current as the motor decelerates.

5. At instant *5* the motor speed is zero and converter unit *2* is operating as a rectifier to supply the motor armature resistance drop. Further advance in the firing of ignitron converter *2* will cause it to continue to operate as a rectifier and to accelerate the motor in the reverse direction during interval *5* to *6*.

6. At instant *6* the motor is running in the reverse direction and is supplied by ignitron unit *2* operating as a rectifier.

7. At instant *7* the load is applied in the reverse direction.

This arrangement has the disadvantage of requiring some duplicate converter equipment. Each ignitron converter unit operates at an average duty cycle of one half of that of the drive motor. Since the peak motor loading contributes in determining the required number and size of ignitrons, this arrangement will in many applications require a greater number of ignitrons than that of other methods which use a single ignitron converter.

An alternate circuit arrangement and operation of this system is to omit the reactors in each converter and to release only one converter at a time so as to eliminate circulating currents between the ignitron converter units. The reactor to limit ripple can then be a single reactor in the motor branch. The operation sequence of this arrangement is similar to that of Fig. 7 except that when it is desired to decelerate in the forward direction, rectifier unit *1* is blocked completely and unit 2, which was previously blocked, is now released to fire as an inverter. Thus, only one ignitron unit is released at a given instant and no circuit is released between the two ignitron converter units for circulating current to flow. It is necessary that the control for this arrangement be worked out so that only one ignitron can be released at a given instant.

Single Electronic Converter with Reversing Switch in Armature Circuit

To eliminate the disadvantages of having two reverse-connected ignitron converters, it is necessary to provide a means to reverse the voltage on a single converter when it is desired to have regeneration. Fig. 8 shows a typical circuit and operation of an arrangement in which the armature circuit of the drive motor is reversed by means of a reversing contactor. The reactor is used to limit the ripple current to the motor, particularly at high ignition delay angles. This method has been applied extensively in Europe using continuously excited tubes;[1,2] it has also been applied on a small domestic installation using ignitrons.

A typical operation sequence during motor reversal for the arrangement shown in Fig. 8 is as follows:

1. At instant *1* the motor is loaded in the forward direction. Current is conducted from the ignitron converter to the motor through the contacts *F*.

2. At instant *2* the load is removed. The motor is now operating at constant speed with friction load.

3. At instant *3* the reversal cycle is begun by lowering the rectifier voltage by phase delay so that it is a small amount below the motor counter-emf. This results in blocking of the motor current. The contacts can open at this time with no current to interrupt and with a minimum of voltage across them.

4. At instant *4* contacts *F* are opened to disconnect the circuit between the ignitron converter and the motor. During the interval from instants *4* to *6* the motor is coasting with its armature open-circuited.

5. Between instants *4* and *5* the ignitron firing angles are changed from rectification to maximum voltage inversion. This change is completed at instant *5*.

6. At instant *6* the motor contacts *R* are closed in the position for reverse rotation, which connects the circuit so that the motor can regenerate into the ignitron converter.

7. During interval *6* to *7* the ignition angles are advanced as the motor decelerates, so that the motor current is nearly constant and is at about rated current. When the motor speed becomes zero at instant *7*, the converter is operating as a rectifier to supply the armature resistance drop.

8. During interval *7* to *9* further advance in firing will cause the ignitrons to continue to operate as rectifiers, and the motor will accelerate in the reverse direction.

9. At instant *8* the motor is up to full speed in the reverse direction and is running at friction load.

10. At instant *9* load is applied in the reverse direction.

This arrangement has the advantage that only a single ignitron converter is required and its duty cycle is the same as that of the mill motor. The disadvantage of this arrangement is the requirement of the reversing contactor and the large number of operations which result in considerable maintenance.

Single Ignitron Converter for Armature and Means to Reverse Field

Another method using a single ignitron converter is to obtain reversal of motor voltage by means of reversing the field. Fig. 9 shows a typical circuit and operation using this method. To reverse the motor as rapidly as possible it is necessary to overexcite the field in order to cause it to reverse in a minimum of time. This can be accomplished by supplying the field from a rotating exciter *E*, as shown in Fig. 9, or by feeding it from an ignitron converter system using two reverse-connected converters similar to that shown for the motor circuit of Fig. 7. Provision must be included to de-energize the armature current by reducing or blocking the rectifier voltage before de-energizing the field or in event of loss of field. This method has been studied and applied in Europe but not as extensively as the previous two methods.[3-4]

A typical operating sequence during motor reversal of an arrangement using the single ignitron converter circuit of Fig. 9 is as follows:

1. At instant *1* in Fig. 9 the motor is carrying load at rated speed in the forward direction.

2. At instant *2* the work load ends and the motor is running at rated speed and no load.

3. At instant *3* the reversing cycle is begun by reducing the armature voltage by phase control of the ignitron converter. The armature current is blocked when the converter voltage becomes less than the counter-emf.

4. During interval *4* to *5* the field voltage is reversed and is several times its rated value. The field current is forced to zero by regenerative action of the rotating exciter *E*. If a dual ignitron converter system is used on the field circuit, one of the units will operate as an inverter during this interval.

5. During interval *5* to *7* the field is built up in the opposite polarity. During interval *5* to *6* the main converter is inverting to provide the required current and torque to cause the motor to decelerate. At instant *7* the field current is at rated value in the opposite polarity and the field voltage is reduced to the normal magnitude.

6. At instant *6* the motor speed is zero and the converter is operating as a rectifier to supply the armature resistance drop.

7. During interval *6* to *8* the converter operates as a rectifier to accelerate the motor in the reverse direction.

8. At instant *8* the motor is at rated speed in the reverse direction and is without load. The field is at its normal value and the armature current is low.

9. At instant *9* the load is applied with the motor running in the reverse direction.

This arrangement has the advantage of using a single ignitron converter for the armature circuit and thus obtaining the best possible utilization of the tubes. The greatest disadvantage of this arrangement is that this method inherently requires a greater time for motor reversal with a conventional motor than the method using dual converters for the armature circuit. This additional time is required to obtain decay and reversal of field, which has a time constant of the order of several seconds. One or more of the following methods may be used to shorten the motor reversal time:

Overexcite the Field. By applying overvoltage of 5 to 10 times normal field voltage, the time required to reverse the field can be minimized.

Laminate the Motor Frame. The use of a laminated motor frame will reduce the circuits for eddy currents and thereby allow the field flux to be changed faster. This motor construction is more expensive but will result in a much lower field time constant.

Use Higher Armature Current. Operation with higher armature current during deceleration can be used to obtain the required torque while the field current is being built up to the opposite polarity.

Control Methods

General Characteristics

The basic method of control in reversing mill drives is to provide armature-voltage control for motor speeds up to and including base speed and to provide field control for speeds above base speed. The direction of rotation is determined by the polarity of the voltage impressed on the armature. The control is designed or interlocked so that operation below base speed and operation during reversal can occur only with rated field current; this is desirable in order to utilize the maximum motor torque without excessive armature current. The required motor operation is accomplished by means of one or more master controllers by which the operator sets the desired speed and

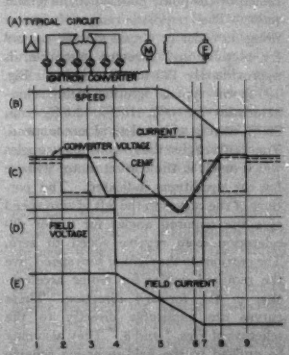


Fig. 9. Circuit and operation of drive with single ignitron converter for armature and means to reverse field

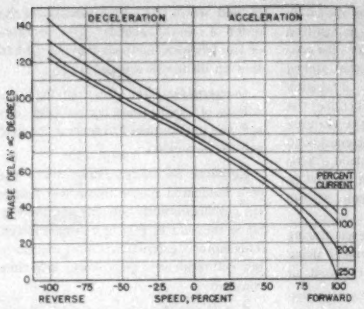


Fig. 10. Ignition angles of power converter for various motor speeds and loads. Motor voltage at base speed and rated current is 80% of converter voltage at no load and zero phase retard. Armature circuit resistance loss is 5% of rated motor input. Field current is at rated value. Converter regulation is 6%. Converter losses are neglected

the desired direction of rotation. The control system then operates to give the proper amount of armature and field control to obtain the desired conditions.

In a conventional type of drive supplied by a motor-generator set the armature-voltage control is obtained by varying the output of the exciter of the d-c machine of the motor-generator set.[8] A rotating regulator or an equivalent static regulator can be used to feed variable voltage of either polarity to the exciter field. Speeds above base motor speed can be obtained by a similar regulator which is required to supply variable voltage of only one polarity to the motor exciter field. The master controller then operates in the control elements of these exciter field regulators.

The method of control in a reversing-mill drive supplied by an ignitron converter system is to obtain voltage variation by phase control of the ignitrons. The ignition time of an ignitron is determined by the phase position of the ignitor pulse. The required range of ignitor firing, in order to obtain voltage control for rectifier and inverter operation, is approximately 120 to 150 degrees. Fig. 10 shows typical ignition delay characteristics for 0%, 100%, 200%, and 250% current as functions of motor speed. These curves give the ignition delay angles which result in the correct motor voltage for the conditions indicated. The portions of the curves to the left of zero speed give the required angles for motor deceleration from reverse speed to zero. The portions to the right of zero speed give the required angles for motor acceleration in the forward direction. These curves are applicable directly to the schemes of Figs. 7 and 8. They also may be applied to the field-reversing scheme of Fig. 9 to obtain ignition angles after the field reversal is completed. The ignition angles for reversal from the forward direction to the reverse direction can be obtained by using opposite polarities for the corresponding points on the speed scale.

Requirements

A control system for this type of application using ignitrons must meet the following general requirements:

1. It must be capable of responding to the various changes in conditions of speed and load. This is necessary as the work being rolled may enter or leave the mill at various speeds as determined by the skill of the operator.

2. It must include means of holding currents to limits determined by ratings of the equipment. This is particularly important during acceleration and deceleration when ignitron equipment is operated with high ignition delay. It is neces.ary that the load not be allowed to exceed the maximum limit during inversion.

3. It must provide for proper reversal time of the main-drive motor so that the motor will be reversed before the work is returned for the next pass. The reversal time requirements should not be made unnecessarily short, as it is probable that an extremely short reversal time will require additional and more expensive equipment, which may make the over-all application uneconomical.

Ignitron Control Methods

The ignitron control system must be capable of continuous variation of ignition angles over the required range within the required reversing time. The method of varying the firing angles is best done by means of static devices because of the short reversal-time requirements. Two general ignitor excitation methods are: 1. reactor firing with phase-shift control by saturable reactors, and 2. thyratron firing with phase-shift control by means of varying the release of the thyratron grids.

Fig. 11 shows the simplified schematic diagram of a typical ignitron excitation circuit using reactor firing. The right side of this diagram shows an extended range reactor firing circuit which supplies the ignitors of two diametrically opposite ignitrons. The feed transformer supplies the operating voltage in the correct phase relationship to the main converter transformer. The secondary of the feed transformer is connected to a static voltage-regulating network through a phase-shifting reactor $X_{PS(1)}$. This network consists of a capacitor C_P and a saturating reactor X_P. The output of this parallel network charges a firing capacitor C_F through a second phase-shift reactor $X_{PS(2)}$. The capacitor voltage is continuously applied to the saturating reactor X_S until the current reaches a value at which the reactor X_S saturates. At this point the capacitor C_F is allowed to discharge through the reactor X_S and a high peak current passes through the ignitor. This current initiates the cathode spot in the ignitron. The dry-plate rectifiers separate the positive and negative pulses so that one firing circuit supplies ignition to two ignitors displaced 180 degrees.

The d-c windings of the phase-shift reactors $X_{PS(1)}$ and $X_{PS(2)}$ are connected in series to a common d-c supply. By increasing the phase-shift current from zero to the maximum value, the ignition angles can be advanced the required range from inverter to rectifier operation. A magnetic amplifier, or Magamp, can be used to control the phase-shift bias power supply which feeds all of the phase-shift reactors of a given ignitron converter. One control winding of the Magamp is adjusted to correspond to the desired speed and direction of rotation; this can be done by proper position of the master controller. Other bias windings on the Magamp provide proper signals and intelligence to hold motor currents to within the respective limits for rectifier and

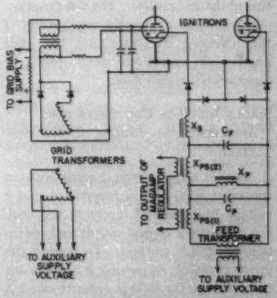


Fig. 11. Typical ignitron excitation circuit

inverter operation. If the reversing-drive system of Fig. 9 with field reversal is used, provision must be included to have the converter voltage less in magnitude than the counter-emf during the interval in which the field current is being reduced to zero prior to its reversal. This method of firing control and regulator operation has advantages of containing static components and of having a sufficiently high rate of response for this type application.

A second method of ignitor firing uses thyratron firing tubes and either a Magamp or an electronic regulator to control the grids. This circuit is advantageous in that the rate of response can be made extremely high, particularly if an electronic regulator is used. The significant disadvantage is the requirement of thyratron tubes, which are replaceable items. It is believed that the tube replacement requirement will prevent the use of this circuit in domestic applications.

With either of the methods of ignitor firing described in the foregoing, the ignitron grids are picked up after the ignitor is fired. The purposes of these grids are to assist pickup of the anode when the ignitor is fired and to assist deionization after the anode conduction has ended. A typical ignitron grid circuit is shown on the left side in Fig. 11. This circuit is for a double-grid ignitron and has the outer grid supplied from two grid transformers connected to different phases of the auxiliary power. These voltages are rectified using metallic rectifiers so as to provide for pickup and operation over the entire ignition range expected. The inner grid is negatively biased except when released by coupling from the outer grid circuit. If a single-grid ignitron is used, a similar grid circuit can be used in which the single grid is released by coupling from the ignitor or an auxiliary anode circuit. This type of grid circuit, in which a grid is negatively biased before conduction is initiated by the ignitor, is desirable in this type of application so as to obtain the maximum ignitron rating under the operation with large amounts of phase control.

Protection

Types of Faults

The faults involved in electronic power converters for this type of application are the conventional faults which appear in power rectifier and inverter circuits. The arc back, which is mainly a rectifier fault, is the failure of the tube to withstand negative voltage and its consequent conduction of current in the reverse direction. Arc backs seldom occur in inverter applications but when they do occur they generally cause arc throughs, which are the usual inverter faults. An arc through is defined as any conduction of forward current in the tube other than during its normal conduction period. One cause of arc throughs is the forward fire, which is the failure of the tube to withstand the forward voltage. A forward fire in rectifier operation under delay only causes the voltage output to be higher for that cycle; in inverter operation the forward fire will generally lead to an arc through. Another cause of inverter arc through is the misfire which results in an arc through of the tube on the preceding phase. A third cause of arc through is commutation failure, which is the failure of the tube, which is ending conduction, to commutate completely its current to to the next phase and to become deionized before its anode-to-cathode voltage becomes positive. Inverter arc throughs are serious in this application because they result in high forward currents and loss of inverter voltage.

In event a fault occurs during the operation of a reversing-mill drive, it is desirable to clear the fault and restore normal operation within a few cycles so as to have minimum effect on the mill-motor operation. An arc back must be cleared from both the a-c and d-c circuits; the a-c circuit can be cleared by electronic arc suppression or by opening the a-c or anode breaker, while the d-c circuit can be cleared only by opening a breaker in the d-c or anode circuit. An inverter arc through must be cleared by opening only the d-c circuit which feeds the inverter.

Protective Devices

The following are the various devices which are used in conventional electronic power converters in various combinations:

A-C Breaker. This can be used for clearing arc backs, for clearing overloads, and for backup protection.

Anode Breaker. This can be used to clear an arc back by opening only the pole carrying the reverse current of the tube that has the arc back. It may be arranged to have all poles trip due to overload or forward fault current during an inverter arc through.

D-C or Cathode Breaker. This is normally designed to trip on reverse current in event of an arc back when anode breakers are not used. It can also be arranged to trip on overcurrent due to an arc through. Two of these breakers can be used in the dual converter circuit of Fig. 7, with one in each converter branch; or a single motor breaker may be applied (with overcurrent trip only).

Electronic Arc Suppression. This can be used to block ignitor firing of ignitrons which feed into faults. This method is not always effective and must be backed up with a breaker. This method can be made most effective when used with the high-voltage type of ignitrons containing more than one grid.

Reactors. These can be placed in various parts of circuits to limit fault current and rate of rise of fault currents. It is undesirable to have too much reactance in the a-c circuit as this makes it more difficult for the inverter to commutate. A d-c reactor can be used to limit the rate of build-up of inverter fault current. Another requirement of the d-c reactor is to limit the current ripple in the d-c machine, particularly in applications having less than six phases. In circuits using reverse-connected converters, such as those of Fig. 7, reactors are required to limit instantaneous circulating currents if both converters are always released.

Protection Schemes

The following briefly describes two general schemes which can be used in all circuits for these applications:

A-C Breaker, Arc Suppression, and D-C Breaker. In event of an arc back on the rectifier, overcurrents cause the rectifier to be blocked by arc suppression, and the cathode breaker is opened by the reverse-current trip. The a-c breaker provides backup on overcurrent in event arc suppression fails. An inverter arc through can be detected by overcurrent in the d-c circuit and can be cleared by opening the d-c breaker. It is unnecessary to arc suppress or open the inverter circuit as its a-c circuit cannot supply a sustained fault current. If the dual-converter circuit of Fig. 7 is used with a single d-c breaker in the motor circuit, an inverter arc through can be cleared by arc suppression of the unit which is rectifying and by opening the motor breaker.

A-C Breaker, Anode Breaker, and D-C Breaker. Rectifier arc backs can be cleared by means of a high-speed anode breaker, which opens only the faulted phase thus minimizing interruption of service. The a-c and d-c breakers can be slower breakers for backup protection. An inverter arc through can be cleared by opening all poles of the anode breaker and using the other breakers for backup protection.

The selection of the basic protection

scheme is dependent upon the characteristics and requirements of the application being considered. Some of the factors which must be studied are the over-all circuit, the number and arrangement of ignitrons, the characteristic of the ignitrons with respect to arc suppression, and economic consideration of the costs of the various protective devices.

In order to minimize interruption of service following a fault, the protection scheme selected should be arranged to provide for fast reclosing once the fault has cleared. Provision should be included for locking out in event the faults are repetitive during a short period. With proper design and application of the ignitron converter equipment faults will seldom occur; and with fast clearing and reclosing, a fault will produce negligible effect on the mill-motor operation.

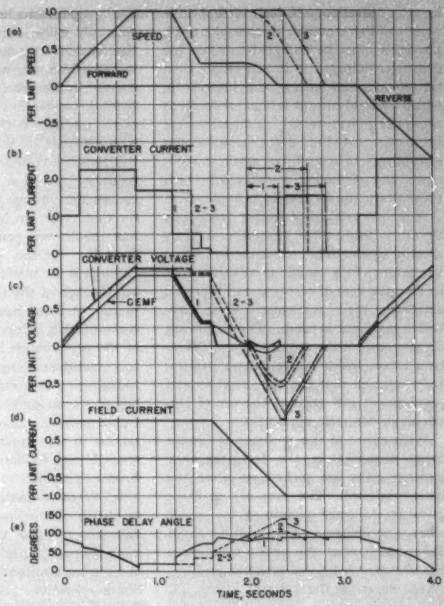


Fig. 12. Characteristics of typical reversing-mill load cycle using single converter with field reversal. Three methods of deceleration are illustrated:

1—Most of deceleration under load
2—All deceleration without load and starting with zero field current
3—All deceleration without load and starting after field current is reversed

Ignitron Application Problems

Description of Reversing-Mill Loading

Fig. 12 shows a calculated cycle of operation during a pass on a steel-mill-reversing drive supplied by an electronic power converter. The speed, voltage, converter current, field current, and ignition delay angles are shown for various time values. The characteristics were determined by analysis based on typical characteristics and requirements of present reversing-mill drives. These characteristics are simplified by omission of transient effects at the changes in operation conditions. The work load is assumed to enter the mill at a low motor speed, at time of 0.2 second. The acceleration of the motor continues with load until time of 0.8 second, after which the rolling operation continues at the desired speed, in this case being base speed. Condition *1*, with solid lines, shows most of the deceleration with load in the rolls. Conditions *2* and *3*, with dashed lines, show two methods of operation in which all of the deceleration occurs after the work has left the rolls.

The deceleration for condition *1* begins at the time of 1.2 seconds, when the firing is delayed to reduce the converter voltage during rectifier operation. This operation is initiated when the end of the work approaches a point near the rolls. The deceleration with load in the rolls continues until the work has left the rolls at time of 1.5 seconds. Further deceleration from 30% speed to zero must be accomplished by means of regeneration by operating the ignitron converter as an inverter, using one of the methods described previously in Figs. 7 through 9.

The method shown in Fig. 12 is that using a single ignitron converter and motor-field reversal. In this sequence the motor remains at the speed at which the work left the rolls until the field current is reduced to zero, which is at time of 2.0 seconds. Deceleration and motor reversal then continue as the field current is built up in the opposite polarity. The current is regulated during decleration; and the converter voltage is near zero. There is little actual regeneration as most of the mechanical power is dissipated in the armature circuit resistance. The maximum ignition delay angle is slightly less than 90 degrees for this condition.

The deceleration for condition *2* begins at the same point as that for condition *1*, but the speed is initially at 100%. The deceleration is longer, and there is more inverter voltage, as a greater amount of regeneration is required. The deceleration after time of 2.4 seconds is with rated field current and is at a constant rate. The maximum ignition delay is about 105 degrees for this condition.

With condition *3* the deceleration is assumed to begin at time of 2.4 seconds with the field current completely reversed. The speed characteristic is linear, since the torque is constant. The motor deceleration time is shorter than that of condition *2*, but the total time measured from the initiation of the field current decay at time of 1.6 seconds is longer. The maximum instantaneous inverter voltage is about 1.8 times that in condition 2. The maximum delay angle is about 140 degrees. This method is less desirable than that of condition *2* for the conditions of Fig. 12 because of the greater total time to decelerate and because of the greater amount of ignition delay required of the excitation and regulating circuits. The deceleration characteristic of condition *3* also is typical for the operation of the circuits of Figs. 7 and 8.

The acceleration characteristics assumed for Fig. 12 indicate that the motor accelerates without load and with rated current from zero to 0.3 per-unit speed in 0.2 second. From this it can be determined that the reversal time from base speed in one direction to base speed in the other direction without load and with rated armature and field currents is 1.33 seconds. The total reversal time for condition *2* of Fig. 12, using an electronic converter with field reversal, is 1.72 seconds, which consists of 0.4 second for decreasing the field current to zero, 0.65

second to decelerate the motor from base speed to zero, and 0.07 second to accelerate from zero to base speed with rated current. This operation is based on an assumed field-reversal time of 0.8 second which may be obtained by overexcitation to several times normal field voltage in a standard motor; this time can be reduced by using a motor with a laminated frame.

The acceleration for the next pass is shown starting at time of 3.2 seconds in Fig. 12. This is 1.7 seconds after the work was removed under condition *1*. In actual operation this interval can be made shorter or longer as determined by the time required for the metal to be returned to the rolls.

Current Requirements

The current sequence in Fig. 12(B) shows that the load current for this type of application consists of a succession of overloads with light load or no load in the period between these overloads. The heaviest loading occurs when the rolling load is first applied and during the acceleration with load; most of this loading occurs at rectifier operation with considerable phase control. The portion of the sequence having inverter operation is quite short and is a relatively small part of the over-all cycle; however, even for this short period of inversion it is necessary to utilize an ignitron design having the proper characteristics. A careful analysis of the worst possible operating conditions can be made to determine the maximum inverter loading with respect to currents and delay angles. If the mill is designed so that condition *1* in Fig. 12 is typical of the maximum phase delay and that operation with higher delay such as condition *2* can never occur, then application can be made with an ignitron design capable of only a small amount of inverter operation. Such a design would be more economical than one that can meet the inverter requirements of conditions *2* or *3* of Fig. 12.

Although National Electrical Manufacturers Association standards have been developed for electrochemical, industrial, railway, and mining applications[6] such standards are not in existence for the various types of steel-mill main-drive service. At present work is being done to develop National Electrical Manufacturers Association standards for continuous-type main drives such as continuous strip and rod mills.[7] Because of the special nature of the loading cycle in reversing-mill drives it is necessary to study each application to determine the number, size, and type of ignitrons required for a given motor rating.[8–10]

Voltage Considerations

The present practice in the United States on conventional rolling-mill drives is to use motors with a maximum rating of 750 volts. This limit has been imposed mainly because of increased motor design problems, such as commutation, flashover at the commutator, and design problems which are inherent at the higher voltages. In addition, the users generally seem to prefer voltages of less than 750 volts for the motors and control equipment. Steel-mill-drive applications in Europe have for many years used some motor voltages of more than 1,200 volts; in the last 10 years there have been applications with motor voltages in the range from 1,200 to 1,600 volts and with use of continuously excited mercury-arc converters.

The application of ignitron tubes to reversing-mill-drive applications requires careful consideration of inverter problems. An ignitron design suitable to withstand forward voltages reliably and to have fast deionization characteristics is best designed with two or more grids.[11] A tube with double grids for inverter requirements inherently has a lower allowable current density than that of a single-grid tube of about the same physical size. Therefore, it is desirable to apply the double-grid type of tube at rectifier voltages above 1,200 volts in order to obtain the optimum kilowatt output for a given size of tube. Although a double-grid ignitron is better for most applications requiring inversion, a single-grid tube is suitable for 750-volt applications requiring only a limited amount of inversion. It may be possible to apply low-voltage single-grid ignitrons to applications in which the inverter countervoltage and delay angles are between those of conditions *1* and *2* in Fig. 12.

If the inverter requirements prevent the use of low-voltage single-grid ignitrons, it will be necessary to use double-grid ignitrons, which can be applied more economically at higher motor voltages, as indicated previously.

It appears, therefore, that consideration should be given in the United States to the use of higher voltage reversing-mill-drive motors in order to more effectively obtain the advantages of using ignitrons for the main drives. Since most present rolling-mill drives have more than one motor, such as a separate drive motor on the individual rolls or a twin motor on the same shaft, one possibility is to connect two lower voltage motors in series to obtain the desired ignition converter voltage. This will still require full-voltage insulation to ground on both motors, however.

Summary

1. Circuits are available to obtain the advantages of ignitron equipment for reversing-mill drives. These circuits use known rectifier and inverter principles and can be controlled so as to provide the proper loading cycle to suit this type of application.

2. Any application using an electronic power converter for a reversing-mill drive must have a power system capable of withstanding the instantaneous power variations of the load. In considering this type of application it is necessary to make careful study of the effects of the expected load pulses on the power system.

3. The dual-converter circuit inherently has the shortest possible reversal time but has the disadvantage of requiring some duplicate equipment. This arrangement generally does not utilize converter equipment to the greatest advantage, especially on the peak loading requirement.

4. The circuit with a single converter and armature reversal using a reversing contactor has the advantages of a single ignitron converter but it is undesirable because of the requirement of the contactor with its large number of operations and consequently large maintenance.

5. The circuit with a single converter and field reversal appears to be the best circuit for application of ignitron equipment. This arrangement may require a motor with a laminated frame so as to obtain a low field time constant. Consideration should be given to utilizing a greater reversal time where the applications will permit such in order to be able to use a motor with a standard frame and a minimum of field forcing.

6. The number, size, and type of ignitrons required for a given motor rating are determined by the load requirements both during rectifier operation at high angles of retard and during inverter operation. To obtain the most economical utilization of ignitrons for inverter service, it is desirable for most applications to have a motor voltage above 1,200 volts. Although this will increase the motor design problems and motor cost, it will provide a more economical power converter cost, since fewer ignitrons will be required for a given power output.

Conclusions

This analysis indicates that ignitron equipment can be applied to reversing-mill-drive applications to obtain the usual advantages of ignitron rectifier equipment. However, further investigation will be required before actual applications can be made. In order to have an economical arrangement, consideration must be given to the selection of equip-

ment and method of operation for a given application so as to obtain minimum overall cost. The use of a mill motor voltage higher than that used at present will increase the attractiveness of the application of ignitron equipment. However, a higher voltage motor will be more difficult to design and will be more expensive both for initial cost and for maintenance. Therefore, in determining the suitability of an ignitron converter for a reversing-mill drive, all factors should be carefully analyzed.

Appendix. List of Symbols

E_a = average direct voltage drop due to loss in the converter elements (average arc drop)

E_d = average direct voltage under load

E_{d0} = theoretical direct voltage (average direct voltage at no load assuming no overlap, no phase control, and zero arc drop)

E_r = average direct voltage drop due to resistance losses

E_s = transformer d-c winding line-to-neutral voltage, rms volts

E_x = average direct voltage drop caused by commutating reactance

I_c = direct current commutated between two rectifying elements in a commutating group

I_d = average d-c load current

p = number of phases in a simple rectifier

u = commutating angle (angle of overlap)

X_c = commutating reactance line to neutral, ohms

α = angle of retard (delay angle, ignition angle), i.e., the angle by which the start of conduction lags the point at which the incoming and outgoing rectifying elements have equal positive voltages

$\cos \alpha$ = phase-control ratio

β = angle of advance, i.e., the angle by which the start of conduction leads the point at which the incoming and outgoing rectifying elements have equal negative voltages

γ = margin angle, i.e., the angle by which the end of commutation leads the point at which the incoming and outgoing rectifying elements have equal negative voltages

κ = conduction angle for one rectifying element.

References

1. Stromrichter in Kreuzschaltung und Gegenparallelschaltung (Current Rectifiers in Cross-Connection Circuit and Counter Parallel Circuit), G. Möltgen. *Siemens-Zeitschrift*, Erlangen, Germany, Dec. 1956.

2. Mutator-Fed Variable Speed Drives, H. Blattner. *Brown Boveri Review*, Baden, Switzerland, Apr.-May, 1955.

3. Stromrichtranlagen für Walzwerkantriebe (Rectifier Installations for Rolling-Mill Drives), F. Hölters. *AEG Mitteilungen*, Berlin-Grunewald, Germany, vol. 44, Sept.-Oct. 1954.

4. Reversierriebten bei Stromrichtergespeisten Umkehrantrieben (Reversing Times with Current-Rectifier-Fed Reversing Drives), G. Möltgen. *Siemens-Zeitschrift*, vol. 28, Nov. 1954.

5. Fairless Works—Electric Equipment for Slabbing Mill and Blooming Mill, R. H. Wright, N. L. Kincaid. *AIEE Transactions*, vol. 73, pt. II, July 1954, pp. 141–44.

6. Mercury-Arc Power Rectifier Units. *Publication no. 220-1955*, National Electrical Manufacturers Association, New York, N. Y., 1955.

7. Considerations in Applying Rectifiers As a Power Supply for Hot Strip Mills, G. M. Zins, E. J. Cham. *AIEE Transactions*, vol. 73, pt. II, May 1954, pp. 65–70.

8. Survey of Operations of Mercury-Arc Rectifiers, AIEE Committee Report. *AIEE Special Publication S-47*, Apr. 1952.

9. Sealed Ignitron Principles Extended to Larger Tubes, J. L. Boyer, A. P. Colaiaco. *AIEE Transactions*, vol 75, pt. II, July 1956, pp. 125–29.

10. Service Life of Pumpless Ignitrons, E. J. Remschied. *AIEE CP 56-174* (available on request).

11. High-Voltage Ignitron Rectifiers and Inverters for Railroad Service, J. L. Boyer, C. G. Hagensick. *AIEE Transactions* (*Electrical Engineering*), vol. 65, July 1946, pp. 463–70.

Discussion

R. V. Pohl (General Electric Company, Schenectady, N. Y.): I agree with Mr. Cham and Mr. Hagensick that the reversing ignitron drive has a bright future.

There are several points in the presentation on which I wish to elaborate. The first is the comparison of dual-ignitron drive versus armature reversal versus field reversal. With the dual-ignitron drive the motor can be under control at all times. With either field reversal or armature reversal, there is a finite dead time when the motor is not under control.

The control system illustrated in Fig. 7 does not show the dual-ignitron drive to its best advantage because it involves a short dead zone. A circuit could be built which would make full use of the advantage of the dual-drive system where the high performance and close control are necessary.

The magnetic firing circuit certainly does have an advantage over the tube firing circuit from the standpoint of life. However, some applications for dual-ignitron drive require the speed that can only be provided by a tube firing circuit. Our field experience has shown that the thyratrons in firing circuit service can provide service for up to 3 years. Where the ultimate in performance is required, tube firing circuits are practical. In these installations, the further advantages of rapid arc suppression and low control power are realized.

Without a doubt, the three types of reversible ignitron drives will find their places. The use of these equipments will increase as higher performance standards are required in the steel-working industry.

W. K. Haessler (Brown Boveri Corporation, New York, N. Y.): The generally known fact that power surges caused by load changes applied to a rectifier-fed mill motor are transferred without attenuation to the supply system has been mentioned in the paper. Since this is a major factor to be considered in the application of rectifiers, we would like to discuss a few important facts regarding the conditions anticipated as a result of load surges, and point to available proved means of how to improve them.

The power factor in the range between zero speed and 100% base speed is rather poor for a rectifier-fed drive. The most disturbing fact is, however, that this results in heavy reactive surges when applying sudden loads to the motor. This condition can in certain networks cause more disturbance than the active power surge due to subsequent voltage drops, thereby considerably affecting the consumer supplied from the same system. It is, therefore, evident that careful investigation, regarding the adequacy of the network, prior to the application of rectifiers, should be made.

The characteristic of the reactive power surges as a function of the motor speed anticipated with a reversing drive operating at 100% overload is plotted in Fig. 13. This characteristic shows that when operating the mill during the initial passes, e.g., between 30% to 50% of base speed (range *A*), the reactive power surges amount to

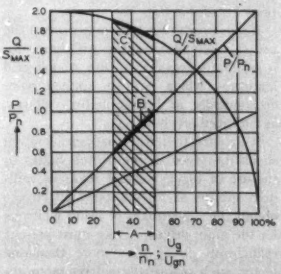


Fig. 13. Reactive and active power of a mercury-arc rectifier with symmetrical grid control at 100% overload

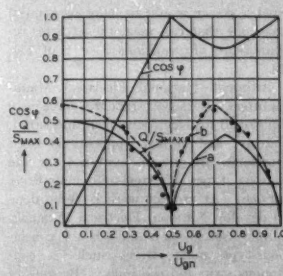


Fig. 14. Power factor (cos φ) and reactive power Q of a mercury-arc rectifier with asymmetrical grid control

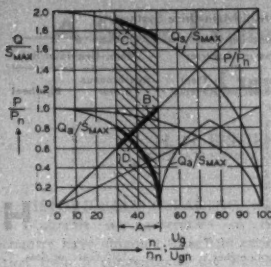


Fig. 15. Comparison of reactive power curves with symmetrical and asymmetrical grid control: Q_s = reactive power with symmetrical grid control; Q_a = reactive power with asymmetrical grid control

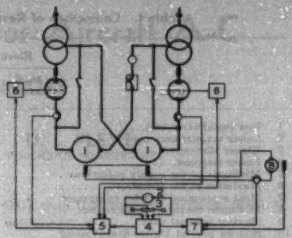


Fig. 16. Principal diagram of rectifier-fed reversing mill with push-pull connection

1—Motor
2—Tachometer dynamo
3—Master controller
4—Filter
5—Armature control
6—Grid control
7—Field control
8—Exciter

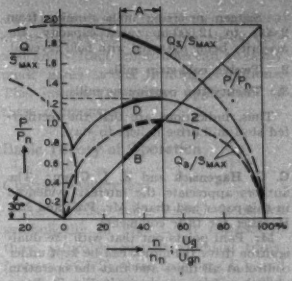


Fig. 17. Comparison of reactive power curves with single-tank connection (Q_s/S_{max}) and push-pull connection (Q_a/S_{max})

S_{max} = maximum apparent power at rated load
P = active power
P_n = rated output
Q_s = reactive power with single-tank connection
Q_a = reactive power with push-pull connection
U_g = direct voltage
U_{gn} = rated voltage at full phase control
n = speed
n_n = rated speed

approximately 180% of the rated active power of the drive. The resulting impact on the power system is considerable especially for high-capacity reversing mills which can cause substantial voltage drops along the reactances of the network. It is, therefore, a basic requirement to connect the rectifier feeder to the strongest point of the available system.

The successful application of rectifiers in plants supplied by low-capacity systems made it necessary to develop means to reduce these heavy reactive power surges. For this purpose, Brown Boveri adopts the so-called asymmetrical grid-control method which has been applied in most of their recent installations using mercury-arc rectifiers. Fig. 14 gives an indication of the conditions obtained with this method. This comprises two additional interphase transformers each having approximately 17% of the rated capacity of the main transformer, a special grid-control set adapted to asymmetrical grid control, and a small 25-cycle motor-generator set.

The measured values of the reactive power as a function of the motor voltage and speed show that the power surges are reduced to about 60% at zero speed whereas at 50% basis speed they practically disappear completely, as the power factor reaches unity. Fig. 15 shows the comparison of the reactive power conditions obtained with these two methods. The average power surges to be expected with asymmetrical grid control are considerably lower over the whole speed range than those obtained with the conventional grid control. Various tests made in existing plants reveal that with this method average reactive power surges are reduced by about 40%.

Another possibility to improve the power factor and thus to reduce the reactive power surges is to connect two sets of rectifiers in series with the reversing motors in the so-called push-pull connection, as shown in Fig. 16. At zero speed (zero voltage), one set of rectifiers works with full voltage as converter and the other set with full voltage as inverter, both having unity power factor. At half speed, one set of rectifiers is working with full voltage as converter and the second set at zero voltage, the resulting power factor being 0.7. At full voltage (full speed), both rectifier groups are operating with full voltage as converters.

As each of the two rectifier groups has to be laid out for the maximum current, this solution is only economical if tandem motors or double-armature motors are utilized and connected in series as shown. This permits the increase of the total direct voltage and the number of rectifiers may be reduced to the original number required for the standard single-tank connection. The resulting reactive power curves are shown in Fig. 17.

With this method, the average peak loads can also be reduced by about 40%.

The great advantage of this control lies in its simplicity, as it needs no special grid control and no special equipment.

It should be mentioned that although it is possible to improve the power factor during normal running by using static capacitators, this will not eliminate the impact of the reactive power surges. Their application with reversing mills is, therefore, not advisable. However, the application of synchronous machines hooked up to the high-tension rectifier bus may, under certain conditions, help to attenuate the impact of the reactive surges, thus minimizing the voltage drop experienced without compensation. The heavy capital cost of such special condensers deterred industry from applying them to any great extent.

In summary, it can be stated that the asymmetrical grid control has proved to be the most economical and convenient way of improving system conditions regarding magnitude of reactive surges. The fact that most of the rectifier-fed drives delivered by Brown Boveri utilizes asymmetrical grid control is certainly a testimony of the usefulness and reliability of this method. Another method with similar results consists in applying the push-pull connection. It should be understood that these methods applied to improve power factor and reactive power surges can be utilized with single as well as multianode units.

A further advantage of rectifier-fed reversing mills should also receive consideration. It is the possibility of applying several rectifiers in parallel as shown in Fig. 18. For example, a 7,000-hp reversing drive can be fed by 14 units of this type (the number of units may vary depending on the required direct voltage) and the failure of one unit would hardly effect the operation of the mill. Furthermore, the replacement of a unit can be made within a short time if a spare unit is available. Since the same type of rectifier can be used in reversing, continuous, and strip mills, it follows that a more than ample stock of spares involves, therefore, a relatively low investment.

During the past 5 to 7 years, Brown Boveri has delivered a considerable number of rectifier-fed steel-mill derives, namely:

Fig. 18. Installation of 6-anode mercury-arc rectifiers in parallel connection supplying steel-mill drive

1. Sixteen reversing mills ranging from 2,000- to 12,000-hp rated capacity, the majority being operated with field reversal.

2. Twelve cold-strip mills.

3. Twenty-two continuous mills.

Thus, indications are that the rectifier-fed steel-mill drive is here to stay.

C. G. Hagensick and **E. J. Cham:** The authors appreciate the interest evidenced in this paper and thank Mr. Pohl and Mr. Haessler for their comments.

Mr. Pohl points out that with the dual-ignitron drive the motor can be kept under control at all times and that the operation of such a drive, as shown in Fig. 7, shows a short dead zone and thus does not show this drive to its best advantage. The authors agree that the dead times during which the motor currents are blocked can be made shorter in the dual-converter drive than with the other arrangements. The interval between instants *3* and *4* in Fig. 7 shows the motor current blocked during the sequence of transferring the motor current from ignitron converter *1* to converter *2*. The length of this interval, as shown in Fig. 7, is not an indication of its relative time value, as explained in the first paragraph under "Power Circuits for Reversing Drives." This interval can be made extremely short, of the order of a few cycles or less of the supply frequency, as it is necessary to change the ignition delay angles of each converter by only a few degrees to effect the transfer of current from one unit to the other.

Mr. Pohl also points out the advantages of thyratron firing for this type of application in order to have extremely rapid control, low control power, and rapid arc suppression. The authors also recognize that the advantages of extremely fast voltage regulation and control available with electronic circuits may overcome the disadvantages of having replaceable tubes in these applications. The advantage of tube firing circuits is further enhanced since the quality and performance of present thyratrons are better than those used more than 15 years ago for the excitation of early ignitron rectifiers, in which their replacement frequency was a major drawback to their use. However, it is felt that the proper design and application of reactor (or magnetic) type of excitation with Magamp regulating systems, as described under "Control Methods," will provide adequate speeds of response for reversing-mill-drive applications without containing replaceable components. In such an arrangement, arc suppression can be obtained by means of high-speed relays, which can cause blocking of the ignitors within a cycle after the fault has occurred.

The discussion by Mr. Haessler describes methods to reduce the high reactive power peaks on the system to which an electronic power converter reversing drive is connected. The method using asynchronous ignition control is also described in reference 2 of the paper and in references 1 and 2 of this closure. Although these methods are effective in reducing the reactive peaks as indicated by Mr. Haessler's mention of their successful application in Europe, it is evident that the asynchronous control method has the disadvantages of requiring additional transformer equipment and, more particularly, a small synchronous control motor-generator set to provide control voltages of one half of the line frequency. Such an auxiliary supply has attendant problems in keeping the control voltages of half of the line frequency synchronized with the rectifier supply system at all times, especially during system transient disturbances.

Table I. Comparison of Reversing-Mill Main-Drive Systems

Item	Electronic Power Converter			Motor-Generator Set	
	Dual Converter	Armature Reversal	Field Reversal	Synchronous Motor	Flywheel
1. Low installation cost		1		2	
2. Lower maintenance cost		1		2	
3. Higher efficiency		1		2	
4. Faster response		1		2	
5. Minimum power supply pulsations		2		2	1
6. Maximum power factor		3		1	2
7. Most economical rectifier equipment	2	1	1		
8. Minimum contactor maintenance	1	2	1	1	
9. Shortest reversal time	1	2	3	2	
10. Best for present United States motor practice	1	1	2*	1	
11. Optimum power supply voltage	1,200 volts and higher			750 volts or less	

* May use standard motor in some applications where reversal times can be met without laminated frame motor.

The other method using two converters in series in the so-called push-pull connection, although not requiring the source of synchronized control at half line frequency, is economically suited only to applications in which two motors can be connected in series. The power factors which Mr. Haessler indicates as available with this method are somewhat theoretical; e.g., the arrangement at zero speed with one converter at full voltage as a rectifier and the other at full voltage as an inverter cannot have unity power factor in each, since the inverter power factor is reduced due to the required ignition advance, and that of the rectifier will be less than unity because the rectifier must be operated with phase control so as to cancel out the inverter voltage. It is estimated that the actual power factor for this condition would be between 70% and 80%, which is still considerably better than that for the equivalent conditions in the conventional circuits. In order to clarify the differences in terminology between Mr. Haessler's discussion and the paper, it is pointed out that the ASA Standard[2] defines electronic power converter as a device with power transfer in either direction between an a-c and a d-c circuit, a rectifier as a device with power transfer from an a-c to a d-c circuit, and an inverter as a device with power transfer from a d-c to an a-c circuit.

The authors recognize that before a reversing-mill drive can be applied using ignitron equipment it will be necessary to study the effects of the power and reactive power pulsations on the power system involved, as is indicated in the paper. The results of such a study would determine whether additional equipment to reduce the reactive power pulsations was necessary or could be economically justified.

Mr. Haessler also indicates that a typical reversing-mill-drive application may consist of many relatively small rectifier units in parallel. Although the advantages for continuity of service in event of failure of a single unit are apparent, it appears that such a large number of small units and the associated number of small transformer winding groups does not offer the greatest possible economy in equipment. The authors believe that a drive system made up of fewer units of larger individual rating will be more economical and can still provide for continuity of service in event of failure of a rectifier element or a complete unit.

The authors note with interest Mr. Haessler's statement that the majority of 16 reversing mills, ranging from 2,000 to 12,000 hp, are being operated with field reversal. Thus it appears that the advantages of better utilization of rectifier equipment offset the problems associated with the field-reversal method. However, the use of field reversal for most of these applications does not indicate that this method is to be preferred for domestic applications, since other factors such as the actual load cycles and reversing-time requirements may be different.

In order to summarize briefly the information given in the paper on the various types of reversing-mill-drive systems, Table I is included as part of the closure. In Table I a comparison is given of the three types of electronic converter drives with synchronous and flywheel motor-generator sets. The numbers indicate the order of importance of various factors and do not represent relative magnitudes. The selection of the converter power circuit and the motor voltage will be dependent upon careful analysis and consideration of these factors for a given application.

References

1. Stromrichteranlagen zur Speisung Motorischer Antriebe (Current Converter Installations for Current Supply to Motor Drives), **F. Hölters.** *AEG Mitteilungen*, Berlin-Grunewald, Germany, vol. 41, Sept.-Oct. 1951.

2. The Utilization of Mutators in Hot Rolling Mills, **U. Matta.** *Brown Boveri Review*, Baden, Switzerland, vol. 43, no. 11, Nov. 1956.

3. Pool Cathode Mercury-Arc Power Converters, *ASA C34.1*, American Standards Association, New York, N. Y., 1949.

A Static Exciter for Aircraft A-C Generators

H. H. BRITTEN
ASSOCIATE MEMBER AIEE

D. L. PLETTE
ASSOCIATE MEMBER AIEE

HISTORICALLY SPEAKING, most aircraft a-c generators in this country have obtained their field excitation from a conventional d-c rotating exciter. This type of exciter is characterized by rotating windings and commutators to produce direct current. Slip rings are used to carry excitation current to the rotating field of the a-c generator.

Recently, another type of rotating exciter has been introduced in which the voltages generated in a rotating a-c exciter are rectified to produce direct-current by small rectifiers which are also rotating. The direct current thus obtained is supplied directly to the rotating field of the a-c generator.

A static or nonrotating type of exciter has also been developed for aircraft a-c generators. This type of exciter extracts a portion of the voltage and current output of the a-c generator and utilizes this power to supply controlled excitation back into the field of the generator. This paper describes a static exciter of this type, the reasons for its use, and some of its characteristics and considerations.

Reasons for Static Excitation

Physical Separation of the A-C Generator and Its Exciter

Due to the fact that this type of exciter is nonrotating, it is not normally attached physically to the a-c generator. It can thus be located at what ever place in the airplane is most convenient from a wiring, installation, and maintenance standpoint. This results in reduced overhung moment on the pad of the engine or constant speed drive on which the generator is mounted. The length of the generator can be appreciably reduced compared with generators with integral exciters. This fact sometimes permits the successful installation of these shorter generators in aircraft where engine accessory space is limited to the extent that rotating exciter generators will not fit.

Effects on Generator Design

The use of a static exciter results in an a-c generator which is simpler in many respects. Many of the rotating parts inherent in generators with integral, rotating exciters are eliminated. These include rotating exciter windings and commutators or rectifiers. A shorter generator shaft and less total rotor weight result in a simpler shaft construction and easier bearing duty. The generator cooling is simplified since there are fewer heat producing elements which must be reached by the heat transfer medium, whether it be air or oil.

The simplification of the rotating portion of the generator due to these factors makes possible high reliability of the generator.

Effects on Exciter Design

The static exciter is made up of components which are of two types: magnetic components such as reactors and transformers, and silicon rectifiers. These components give long, reliable service provided they are not subjected to temperatures, thermal and mechanical shock, vibration, forward current, and peak inverse voltages, which are in excess of their ratings. The physical design of a static exciter is such that the measurement of the temperature, vibration, shock, current and voltage amplitudes can be easily made under actual limiting conditions. This insures the proper application of these components.

Basic Theory of Operation

Circuit Description

The 3-phase static exciter is shown symbolically in Fig. 1. It consists of three saturable current-potential transformers (SCPT), three reactors (X_L), and a 3-phase full-wave rectifier. The SCPT's have three a-c windings and a d-c control winding. The a-c windings are designated N_1, N_2, and N_3, and are respectively the voltage, current, and output windings. The output windings are connected in delta.

The SCPT's combine voltage and current quantities from the generator output in the correct proportion and control the magnitude of the resultant field current in the generator by means of d-c saturation. The SCPT's also act as an impedance matching member between the generator output and the generator field. The reactors shift the phase of the current derived from the generator voltage with respect to that derived from load current such that the field current, which is proportional to the vector sum of these currents, has self-corrective action with respect to changes in load and power factor. REC is the main field rectifier and converts the 3-phase output of the exciter to d-c field excitation.

Single Phase Equivalent Circuit

From writing the equations for the steady-state characteristics of the circuit shown in Fig. 1, the single-phase equiv-

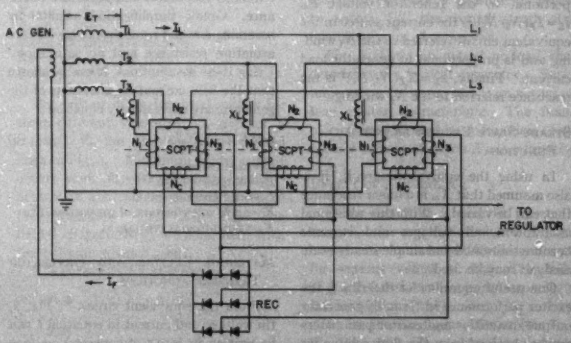


Fig. 1. Symbolic representation of static exciter

Paper 58-213, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted May 16, 1957; made available for printing November 26, 1957.

H. H. Britten and D. L. Plette are with the General Electric Company, Waynesboro, Va.

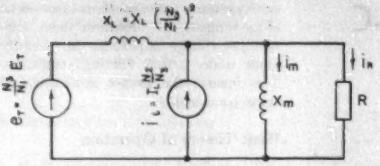
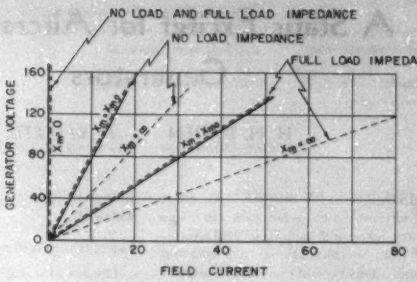


Fig. 2 (above). Single-phase equivalent circuit

Fig. 3 (right). Idealized generator and exciter transfer characteristics



——— Generator characteristic - - - Exciter characteristic

alent circuit of Fig. 2 can be derived by making the following simplifying assumptions.

1. The generator is symmetrically loaded.
2. The reactors have negligible winding resistance and core loss.
3. Leakage reactance, winding resistance, and core loss of the SCPT are negligible.
4. The field of the generator can be represented in the steady state by a linear resistance in the single-phase equivalent circuit.
5. The direct current in the field is proportional to the alternating current flowing through the linear resistance in the single-phase equivalent circuit.

The equivalent circuit can be referred to any of the three a-c windings of the SCPT. In the circuit of Fig. 2, voltages, currents, and impedances are referred to the output winding N_3. R is the per phase resistance representing field loading on the exciter. X_m is the exciting reactance of the SCPT referred to the N_3 winding and is a function of the d-c control current. $e_T=E_T(N_3/N_1)$ is the voltage source in the equivalent circuit referred to the N_3 winding and is proportional to the generator voltage E_t. $i_L=I_L(N_2/N_3)$ is the current source in the equivalent circuit referred to the N_3 winding and is proportional to generator load current. Finally, $x_L=X_L(N_3/N_1)^2$ is the reactance referred to the N_3 winding.

Steady-State Exciter Transfer Function

In using the equivalent circuit, it is also assumed that X_m is a linear reactance that can be varied. With this additional assumption, all voltages and currents become sinusoidal and simple steady-state analysis may be used.

One useful equation for describing the exciter performance in terms of generator output quantities and exciter parameters can be obtained from Fig. 2 by solving for i_R. This is given as equation 1.

$$\bar{i}_R=\frac{\bar{e}_T+jx_L\bar{i}_L}{R\left(1+\frac{x_L}{x_m}\right)+jx_L} \qquad (1)$$

Equation 1 can also be written as:

$$|\bar{i}_R|=\frac{1}{\sqrt{R^2\left(1+\frac{x_L}{x_m}\right)^2+x_L^2}}\left|\bar{e}_T+jx_L\bar{i}_L\right| \qquad (2)$$

As mentioned previously, it has been assumed that the d-c field current will be proportional to the absolute value of rms current in R. Equation 3 expresses this

$$i_F=K_{1R}|\bar{i}_R| \qquad (3)$$

We get the steady-state exciter transfer function combining equations 2 and 3

$$i_F=\frac{K_1}{\sqrt{R^2\left(1+\frac{x_L}{x_m}\right)^2+x_L^2}}\left|\bar{e}_T+jx_L\bar{i}_L\right| \qquad (4)$$

Steady-State Generator Transfer Function

The field current requirements of a synchronous generator can be expressed in terms of its terminal voltage, load current, direct and quadrature axis synchronous reactance, and armature resistance. Great simplification results by assuming a round rotor machine with zero armature resistance and no saturation.[1] Using these assumptions, it can be shown that the field current requirements of the generator are expressed by equation 5

$$i_F=K\left|\bar{E}_T+jX_d\bar{I}_L\right| \qquad (5)$$

i_F = field current
E_t = output voltage
I_L = load current
K = amp/volt constant of proportionality
X_d = synchronous reactance

Matching Exciter and Generator Transfer Functions

From the equivalent circuit of Fig. 2, the voltage and current in equation 4 can be written in terms of generator voltage and load current as shown in equation 6

$$i_F=\frac{K_1}{\sqrt{R^2\left(1+\frac{x_L}{x_m}\right)^2+x_L^2}}\left|\left(\frac{N_3}{N_1}\right)\bar{E}_T+jx_L\left(\frac{N_3}{N_2}\right)\bar{I}_L\right| \qquad (6)$$

Since in equation 6 there are available two independent choices of turns ratios, it is possible to match the coefficients of equation 6 to 4. When these coefficients have been matched, the output of the exciter (equation 6) as a function of generator voltage and load parameters equals the field current requirements of the generator for the same voltage and load conditions. The idealized exciter, therefore at a fixed field temperature can supply correct excitation to the idealized generator for all symmetrical load conditions without any steady state change from the voltage regulator. Equation 6 also shows that a change in the resistance of the field will cause a mismatch of the exciter and generator. Changing the value of X_m slightly by action of the voltage regulator will restore the matching.

In matching the coefficients of equations 5 to 6, the impedance levels of X_m, x_L, and R should be specified before the turns ratios are chosen. It is convenient to refer x_L and X_m to R since this is a function of generator field resistance.

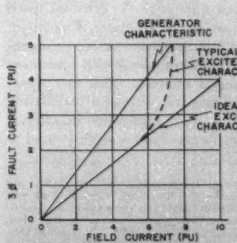


Fig. 4. Typical generator and exciter transfer characteristic during a symmetrical fault

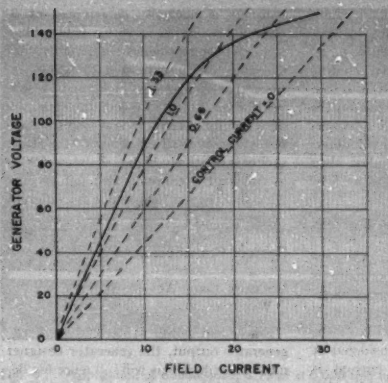
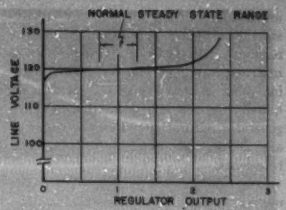


Fig. 5 (left). Typical no-load generator and exciter transfer characteristics

—— Generator characteristic
– – – Exciter characteristic

Fig. 6 (right). Typical steady-state regulator transfer characteristic



This resistance value is normally determined before the design is started.

X_m and x_L can be referred to R at a field temperature of 25 C (degrees centigrade) as follows:

$$X_{m0} = K_m R_0 \tag{7}$$

$$x_L = K_L R_0 \tag{8}$$

R_0 is the value of R at 25 C. X_{m0} is the specific value of the variable X_m which will match the coefficients of equations 5 and 6 with $R = R_0$ and the fixed reactance x_L chosen according to equation 8.

To match the exciter to the generator with the field resistance at 25 C, equations 5 and 6 are used with the values defined in 7 and 8. The resulting SCPT turns ratios are as follows:

$$\frac{N_3}{N_1} = \frac{KR_0}{K_1}\sqrt{\left(1+\frac{K_L}{K_m}\right)^2 + K_L^2} \tag{9}$$

$$\frac{N_2}{N_1} = \frac{KX_d}{K_1 x_L}\sqrt{\left(1+\frac{K_L}{K_m}\right)^2 + K_L^2} \tag{10}$$

Self-Excitation Lines

Self-excitation of a-c generators is relatively new.[2] Self-excitation of d-c generators is a relatively old and well-known method of supplying excitation to shunt wound machines. The common method of presenting the machine data is to plot generated voltage as a function of field current and field current as a function of generated voltage on the same axis. The latter characteristic is known as the field resistance line. Those who have applied carbon pile regulators are familiar with the variation of this line as a function of pole resistance. In the a-c exciter, the field excitation lines are varied statically by means of the variable X_m.

Another method of self-excitation used with d-c machines is series compounding. Here the excitation is a function of generated voltage and load current. The static exciter is a method of self-excitation for a-c machines that closely parallels the series compounded d-c machine in that excitation is a function of voltage and load current. In addition, the output of the a-c exciter is a function of power factor.

As with d-c machines, the characteristics of the a-c exciter may be displayed as self-excitation lines. The output current of the exciter can be plotted as a function of generated voltage alone for any specified fixed load impedance on the generator. A series of field lines can be plotted for the a-c exciter as a function of the variable X_m.

In Fig. 3, the characteristics of an idealized machine at no-load and full-load impedance, rated power factor (pf) are shown as solid lines. The constants of the machine are

$K = 0.125$ ampere/volt
$X_d = 2.43$ ohms
$R_0 = 0.70$ ohm
Rated current = 111 amperes
Rated voltage = 120 volts
Rated pf = 0.75

The output of an idealized exciter for three different values of X_m and for a constant R_0 are also shown as dashed lines on Fig. 3. Maximum exciter output occurs when $X_m = 00$. The exciter characteristics are coincident with those of the generator when $X_m = X_{m0}$. The exciter characteristics are coincident with the ordinate when $X_m = 0$.

The constants for this exciter are

$K_m = 0.8$
$K_L = 3$
$K_1 = 2.08$
$N_3/N_2 = 0.274$
$N_2/N_1 = 0.236$

Performance During a 3-Phase Fault

During a 3-phase fault, the generator transfer function of fault current as a function of field current approaches the idealized generator since it is linear in the range of practical fault currents.

In the exciter, the potential source becomes zero and X_m increases to its maximum value due to regulator action. Thus, neglecting X_m, the exciter equivalent circuit is a current source shunted by a reactance and the field.

Fig. 4 shows the transfer function of a typical generator and an idealized exciter. The dashed line shows the limitation of a practical exciter at 25 C field temperature. The ultimate fault current attained will be determined by the point at which the generator and exciter characteristics intersect. Thus, from Fig. 4 the fault current will be five per unit.

Design Considerations

Nonidealized or Actual Case

Under theory of operation, certain assumptions were made concerning the generator and exciter in order to clearly illustrate the principle involved in the static exciter. Design equations developed under much more rigorous assumptions make it possible to predict to within 5% the actual steady-state operating point of an exciter.

Fig. 5 shows an actual no-load characteristic of a typical generator. Superimposed on the same axis is an actual no-load exciter characteristic. The family of curves is plotted as a function of d-c control current of which X_m is a function.

Effect of Varying K_m

From equations 9 and 10 it can be seen that for a specific value of K_L, the exciter can be matched to the generator for various values of K_m. The value of K_m determines the reserve capacity or forcing ability of the exciter.

In the idealized generator-exciter combination, the exciter is capable of supplying correct excitation to the generator for

any load condition without any steady-state correction required from the voltage regulator. In practice, if an exciter is matched to a generator for normal loading, some reserve capacity is usually required for over-load conditions due to generator saturation. Reserve capacity also allows faster recovery from load transients.

Lowering the value of K_m increases the reserve capacity or forcing ability of the exciter. Since lowering K_m causes an increased exciter weight, the highest value of K_m which will fulfill the requirements for response time and overload capacity is used in normal practice.

Effect of Varying K_L

The influence of K_L on exciter performance can be predicted by inspection of the equivalent circuit of Fig. 2. As K_L becomes larger, the current from the potential source appears to be derived more nearly from a constant current source. Since the other component of field current is derived from a current source, the result is less sensitivity to changes in field resistance as K_L becomes larger.

In practice, K_L is limited in maximum value by the fact that as it is increased the peak inverse voltage impressed on the output rectifiers is also increased. A compromise is finally determined based on the ability of the regular to correct for changes in field resistance and based on availability and cost of higher voltage rectifiers.

Component Design Considerations

The current in the reactors and the potential winding of the SCPT is greater at no-load than full-load rated power factor. Thus, the reactor current rating should be determined at no-load conditions. The volt-second rating (saturation level) of the reactors should be equal to or greater than the potential winding of the SCPT. The reason for this is that during a symmetrical short circuit on the generator, the reactors are in parallel with the potential winding and will determine the excitation ceiling unless they saturate at a higher level than the SCPT.

The current ratings of the SCPT output and current windings are mainly determined by full-load conditions since the thermal time constant normally is such that during relatively short overload periods, the additional temperature rise is small. The current rating of the potential winding must be considered at no-load conditions since its current is greatest at this point. Consideration must be given to core loss at full-load and hot field since core loss is greatest at the highest output voltage condition. The choice of the control winding involves proper matching to the regulator to obtain required steady-state and transient performance. In the practical designs, the watts dissipated in the control winding are negligible compared to the total dissipation at full-load. The volt-second rating (saturation level) of the SCPT must be considered at ceiling excitation conditions which generally occur at three-phase fault or maximum overload requirements with a hot generator field.

The size of the required heat sinks for the silicon rectifiers must be considered at full-load and 150% load. The thermal time constant of the heat sink is such that considerable temperature rise can occur during a 5-minute overload period. The voltage and current rating of the rectifiers is normally determined at 3-phase fault conditions with a cold field.

The silicon rectifiers are inherently very efficient. Because most applications require that components be cooled by self-ventilation, magnetic components are designed for relatively low losses. As a result, the over-all efficiency of the static exciter is relatively high. A value of 90% at full load is common although it may vary slightly from this value for different designs.

System Considerations

Voltage Drop Across SCPT Current Winding

Since field power is derived from the generator output, the generator designer must know how this will influence his design. Field power is derived from both potential and current sources; however, nearly all of this power, under full load and overloads, is obtained from the current source. At no load, all power is obtained from the potential source, which results in a small current being drawn from the generator. Since this current is further decreased as load is added to the generator, its effect can be neglected in considering the design of the generator. The voltage drop across the SCPT current winding is apreciable and must be considered.

It is evident from the equivalent circuit of Fig. 2 that the voltage drop across R is the voltage drop across the SCPT secondary. The voltage drop across the primary of the SCPT is therefore merely the turns ratio times this voltage. From equation 2 and using the relations of 7 and 8, the voltage drop across the SCPT primary is approximately

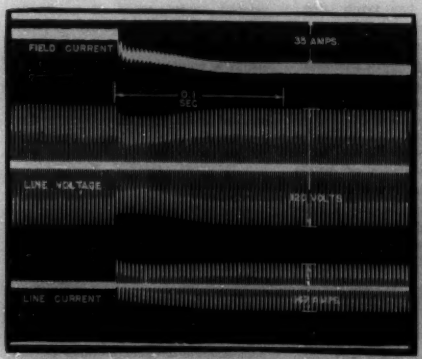


Fig. 7. Application of full load, 0.75 pf at 6,000 rpm to 60-kva generator

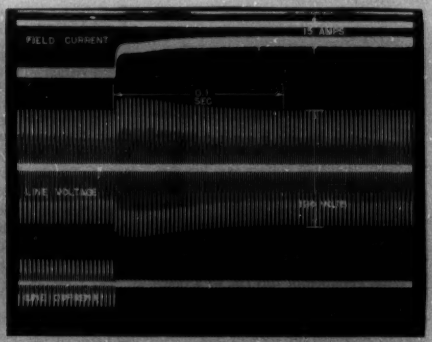


Fig. 8. Removal of full load, 0.75 pf at 6,000 rpm from 60 kva generator

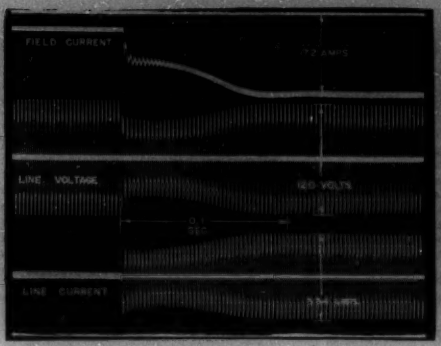


Fig. 9. Application of double-load, 0.75 pf at 6,000 rpm to 60-kva generator

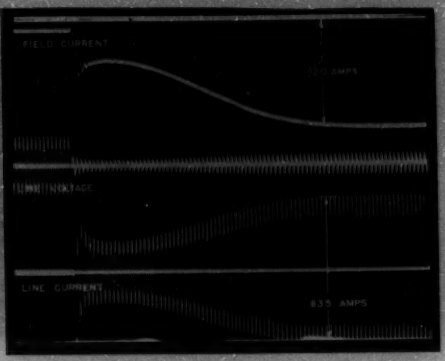


Fig. 10. Application of 3-phase fault at 6,000 rpm to 60-kva generator

$$\bar{V}_P = \frac{\left(\frac{N_2}{N_1}\bar{E}_T + jx_L\frac{N_2}{N_1}\bar{I}_L\right)R\frac{N_2}{N_1}}{R\left(1+\frac{x_L}{x_m}\right)+jx_L} \qquad (11)$$

or

$$\bar{V}_P = \frac{\left(\frac{N_2}{N_1}\bar{E}_T + jx_L\left(\frac{N_2}{N_1}\right)^2\bar{I}_L\right)}{\left(1+\frac{x_L}{x_m}\right)+\left(j\frac{x_L}{R}\right)} \qquad (12)$$

Inspection of equation 12 reveals that even at no load there is a voltage drop across the SCPT primary. This voltage lags the generated voltage an amount determined by the X_L, X_m, and R parameters. As full-load rated power factor is approached, the drop increases and approaches more nearly the phase of the generated voltage.

On a typical 60-kva machine at full-load and hot field conditions, the series drop across the SCPT primary would be approximately 2%.

Effect of Exciter on Waveform of Generator

Since the current drawn by the static exciter is negligible, only the voltage drop across the current winding needs to be considered as far as any contribution to this voltage drop is 1 to 3% of the line-to-neutral voltage, the contribution to harmonics could not be very great. In actual practice the line drop seems to be largely fundamental, and as a result there is only a very small change in the waveform due to the static exciter.

Regulator Requirements

Increasing the value of X_m will increase the output of the exciter at any given load and the converse is also true. X_m increases as the control current in the SCPT decreases. Hence, the sense of the voltage regulator should be such that the regulator output current will increase with increasing generator voltage and be zero at zero voltage. Thus, during a transient caused by load application, the voltage will dip and cause the regulator output to decrease. The exciter output will increase due to load application, and still greater forcing will be obtained as the regulator output decreases. The result will be rapid voltage recovery. A typical steady-state regulator transfer characteristic is shown in Fig. 6.

Effect of a Short Circuit Between the Generator and Exciter

In the event that a short circuit is applied between the generator terminals and the current winding on the SCPT, there will be a tendency for the exciter output to collapse to zero. This tendency is caused by the fact that the short circuit requires additional excitation in the generator field. Since the current flowing into the fault does not flow through the current winding, the exciter does not inherently supply the required excitation to the generator. Providing the fault is of high impedance, the regulator action may have sufficient effect in raising the value of X_m to prevent the system voltage from collapsing. On a low impedance fault, the field excitation and the voltages and currents on all phases will collapse to zero. The exact fault impedance at which the excitation will collapse is a function of the generator reactances, the number of phases faulted, the normal load on the system, and the constants of the exciter.

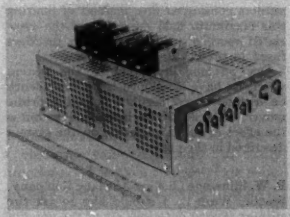


Fig. 11. Static exciter for 60-kva generator

De-energizing the Exciter

In order to de-energize the static exciter system, the SCPT potential windings are shorted. Due to the coupling between the windings of the SCPT, this effectively short-circuits all windings, causing the exciter output to go to zero. This will cause the generator output voltage and current to go to zero as soon as the energy stored in the field inductance is dissipated. Three contacts of a relay (generator control relay) are actually used to short all three potential windings; however, a short across one potential winding is sufficient to de-energize the exciter. The duty of the relay contacts is extremely light since these contacts merely have to "make" a current which is soon reduced to zero. The initial maximum value of the current is equal to the load current times the ratio N_2/N_1. In all designs made to date, this has been a value well within the contact rating of a standard 5-ampere relay. When these

generator control relay contacts open, there is generally no current flowing, and therefore there is no current interruption problem.

Flashing for Initial Build-Up

Since the residual magnetism of the a-c generator is made very small to permit complete de-energizing, a means of flashing is required to initiate the self-excited build-up of voltage. This flashing excitation is usually provided from a small permanent magnet generator. The magnet generator is built as a part of the main a-c generator. In this way, positive build-up is assured.

Results

The following are data taken from a typical 60-kva system at 25 C:

1. Steady-state voltage regulation of ±0.5%, from no load to full load, 0.75 pf.

2. Transient performance as shown in Figs. 7–10.

3. Most harmonics changed less than 0.1% due to the contribution of the exciter. Two harmonics were increased by as much as 0.4% and one harmonic was decreased by 0.5%.

4 A picture of the static exciter is shown in Fig. 11.

Conclusions

A static or nonrotating type of excitation system has been shown to be practical for aircraft a-c generators. This system permits flexibility of installation, produces simplicity in the a-c generator design, and provides excellent reliability and performance.

References

1. Synchronous Machines, Theory and Performance (book), C. Concordia. John Wiley & Sons, Inc., New York, N. Y., pp. 6–31.

2. Static Magnetic Exciter for Synchronous Alternators, H. F. Storm. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 1014–17.

Discussion

H. A. Rubinstein and **G. R. Slemon** (University of Toronto, Toronto, Ont., Canada): This paper is of considerable interest to us since we have developed and tested a somewhat similar excitation system for a-c generators at the University of Toronto. The system which we used differs from that shown in Fig. 1 of the paper in the following aspects:

1. No auxiliary regulator is used and no d-c windings are required on the transformers.

2. The series windings of the transformers are connected on the generator side of the shunt windings to compensate for the current fed into the reactor X_L. This connection assists in the initial build-up and in the recovery from load transients.

3. The reactor X_L was replaced by a capacitor with reversed shunt winding in some of our tests with satisfactory results.

4. No permanent magnet generator was used.

In the design of our system the magnetizing and leakage reactances of the generators were taken into account. Correction was applied for the change in generator magnetizing reactance with change in saturation as well as for the leakage reactance of the transformer windings. Tests were made on a 5-kva 60-cycle alternator fitted with this self-excitation system. The steady-state performance may be summarized as follows:

1. With a reactor in the exciter circuit the voltage regulation from no load to full load was ±1.5% for pf from unity to zero lagging. At 0 pf lagging the voltage regulation was 5% between no load and 200% load.

2. With a capacitor in the exciter circuit the voltage regulation to full load was 1.5% for pf = 1.0 and −6% for pf = 0 lagging. When the load was increased to 270% at pf = 0 lagging the regulation was 1.5%.

3. Using the reactor a 1% change in alternator speed resulted in a 1% change in voltage. With the capacitor a 1% change in speed caused a 3.3% change in voltage.

Reasonably good voltage regulation can therefore be obtained without the use of an auxiliary regulator. This is true even in the case of large inductive overloads such as would occur while starting induction motors of comparable rating at full voltage.

The use of a capacitor in the excitation system is permissible only if good speed regulation is available. The capacitor has however the advantage of improving the load power factor and increasing the net output capability of the alternator while the reactor decreases this capability.

As mentioned previously no permanent magnet generator was found to be necessary to provide initial build-up of voltage. The residual magnetism of approximately 2% was found to be sufficient to cause self-excitation. The time for build-up of voltage after the field circuit was closed was approximately 0.6 second.

When full load was applied the voltage recovered in approximately 0.08 second. The corresponding time of voltage recovery for full load removed was about 0.12 second. There is insufficient information to compare this speed of response with that given in the paper but it would appear that adequate response speed can be obtained without a regulator in some applications.

It was found that dangerously high inverse voltages were produced across the rectifier during load switching transients. We would like to ask whether the authors have had similar experience and if so what methods they would suggest for protecting the rectifiers. Also it appears from Fig. 10 of the paper that sufficient rectifier capacity was installed to provide 120 amperes field current during a short circuit. Have the authors considered any method of limiting this maximum field current and the associated rectifier rating? Are the rectifiers fused to protect against overcurrents?

One obvious limitation of the system which we have developed is its dependence on a good speed regulation in the prime mover. We would be interested to know the effect of speed variation on the system described in the paper.

R. W. Stineman (Boeing Airplane Company, Seattle, Wash.): I should like to ask the authors several questions about the use of a static exciter for an a-c generator. First, it is well known that brushes tend to be troublesome at high altitude and high temperature. The static exciter eliminates the commutator and commutator brushes, but requires slip ring brushes. Can the authors provide any information about slip ring brush performance at high altitude and high temperature? What are the altitude and temperature limitations of the static exciter, as imposed by the brushes?

Another important question concerns weight. Does the static exciter have a weight advantage or penalty compared with a conventional d-c exciter or a brushless a-c exciter? If so, can the authors give a few typical figures? It is realized that there may be installation differences which affect over-all weight, but such factors need to be considered individually for each case. It is also necessary to know what weight increment to associate with the equipment itself, this increment being independent of the installation.

H. H. Britten and **D. L. Plette**: We wish to express our appreciation for the interest shown in the paper by Messrs. Rubenstein and Slemon and also by Mr. Stineman. The type of exciter mentioned by Messrs. Rubenstein and Slemon is interesting in that an attempt has been made to compensate quite exactly for load changes, thereby eliminating the need for a voltage regulator. Our company has also built static exciters for nonaircraft applications which can be operated with or without a voltage regulator. These schemes are somewhat limited, however, when wide variations in field temperatures are present. In the application described in our paper, generator field temperatures ranging from −55 C to +250 C made it desirable to use a voltage regulator.

The permanent magnet auxiliary generator was used in our system for several reasons:

(1) The application required positive build-up characteristics regardless of previous history or machine environment and without dependence upon any external power source.

(2) It was found convenient to use the permanent magnet generator for other purposes such as supply for the voltage regulator and other control equipment.

(3) In the event of a short circuit between

the generator and the bus, it is desirable to remove the excitation from the generator completely to minimize the chances of fire.

High inverse voltages are produced across the rectifiers due to action of the exciter during heavy short circuits and due to voltages induced in the generator field. With the generators which are utilized, the latter effects were less severe and therefore were not a problem. No attempt was made to limit the field current during short circuits even though excess short circuit current is available at lower generator field temperatures. It was determined that the reduction in rectifier capacity requirements did not justify the increased weight and complexity of current limiting circuitry. The rectifiers are not fused and are protected from sustained short circuits in the same manner as the generator by the system undervoltage protection.

The main effects of speed variation on the system described in the paper are reduced overload and steady state load capacity. Effects on regulation are small and can be easily taken care of by the voltage regulator.

Regarding Mr. Stineman's question on the related subject of slip-ring brush performance, slip-ring brush life is averaging more than 2,000 hours on statically excited generators. Slip-ring brush wear tests have been run very successfully with altitudes in excess of 80,000 feet and at temperatures of 400 degrees Fahrenheit. Chemical analysis and laboratory tests on various brush compositions and slip ring materials have given our generator designers confidence that brushes will be satisfactory above 800 degrees Fahrenheit.

Static exciters in aircraft are normally mounted separately from the generators in order to reduce the weight and overhung moment on the engine pad. Since it is usually not desirable to run ducting for forced air cooling to these remote locations, it becomes necessary to design a unit with high efficiencies such that self-cooling can be achieved. This fact does tend to make static exciters of the type we are presently building somewhat heavier than their rotating exciter counterparts which are designed for lower efficiencies. However, we have found that weight comparisons cannot be made on any one system component with any real meaning. For example, the regulators for static exciters are usually appreciably simpler and lighter than those generally employed in rotating exciter systems of the types Mr. Stineman mentions. There are applications considerations which may also result in a weight advantage for either static exciter systems or rotating exciter systems. The authors do not feel, therefore, that any generalization can be made which would be useful

Lighting the Modern Commercial Aircraft

PHILIP E. MASSIE
ASSOCIATE MEMBER AIEE

THE ADVENT of an aircraft wherein 115-volt alternating voltage is the primary power system offers the lighting engineer an opportunity to exploit a wider field of lighting devices. Previously, aircraft could use alternating voltage only at a sacrifice as it was heavy, expensive, and inefficient. However, it now occupies an advantage over 28-volt direct voltage for lighting use. It is the purpose of this paper to show how lighting research and alternating voltage contribute to the lighting of a modern commerical aircraft.

Progress is important in products that must appeal to the public market. Each new transportation vehicle must show distinct advances over previous models. This progress may be present in a variety of ways: eye appeal, performance, comfort, safety, efficiency. In present-day aircraft design, the term "$5,000,000 look" expresses the attempt to increase public acceptance. Add to this appeal the increased operating speeds involved, requiring a higher degree of performance by the crewman, the high cost of weight and time, and the expense of poor aerodynamics. These all contribute to the necessity for progress in design.

Paper 57-458, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted January 3, 1957; made available for printing May 13, 1958.

PHILIP E. MASSIE is with Day-Ray Products, Inc., South Pasadena, Calif.

Contrast and Performance

Performance of motor tasks is usually dependent on the ability to see. This is particularly true of pilotage. The quality of visibility governs the speed and accuracy with which a trained operator can perform a given task. In aircraft pilotage this task is complicated and is compounded by a tremendous responsibility.

The following definitions are given to assure that the reader has the same concept for a term as has the author. These apply to this paper and are generally in common usage where there is agreement on the definition. Brightness ratio will be used to compare the brightness of two areas. The visual field is divided into object, background, and surround. The object is the thing being observed and is assumed to subtend 1/2 degree. The background is the area covered by foveal vision and subtends about 1 to 3 degrees. The surround is the rest of the visual field and covers about 6 steradians.

Brightness ratio as seen by the pilot is a changing value depending on many functions. The fixed functions are cockpit color, maximum interior lighting level, and interior surface reflectance. The variable function is sky brightness. This varies as a function of altitude, heading, time of day, season, and weather.[1-4] Measurements made in a transport cockpit facing the north sky at sea level on a slightly hazy day, in the middle of the morning in Santa Monica, Calif., showed the pilot facing a brightness ratio of 300 in a dark-green cockpit and 175 in a medium-gray cockpit. The only difference between the two cockpits was the color. The reduced brightness ratio in the gray cockpit produced better cockpits by a small margin. The brightness ratio of 175 is still excessively high.

The conditions noted do not provide the maximum brightness ratio. The principal glare source faced by the pilot is the sky. The sky brightness fills an area in front of the pilot of roughly 2 steradians.[4] Assuming a clear day, sky brightness can be expected to decrease as the observer's altitude increases; to decrease as the observation altitude angle increases; to decrease at a heading angle 90 degrees from the sun and vary with the sun's altitude angle. The foregoing measurements showed a brightness of 1,500 to 4,000 foot-lamberts. This would be an extremely bright surface if placed in a room, but may be exceeded out of doors. The pilot is in a room looking out of doors and yet trying to see in the room also. He faces the brightness ratios noted and this results in difficult seeing conditions. Under these vision conditions the pilot operates at reduced speed, efficiency, and accuracy, and the strain may materially affect his judgment and reaction time. An attempt to analyze the visual efficiency of this condition resulted in known empirical formulas going to infinity or beyond the tabulated values.[7,8]

The night vision problem is the reverse of the day problem. The pilot wishes to

observe in a very low illumination level for outside vision. He must still be able to see the multitude of operational devices inside the cockpit and make relatively fine determinations. The outside vision problem may vary from contrast determination to form determination.[9] Vision in this low light level is called scotopic vision as compared to photopic vision for high-light-level vision. Scotopic vision is characterized by reduced sensitivity to red light.[10] Thus the cockpit illuminated by low-level red light results in a lower adaptation level of the pilot's eyes. This enables him to see well both inside and outside the cockpit. Outside vision would be best with no inside illumination. However, this is impossible except for short periods of time. The low-level red light is the best substitute.

High-Altitude Lighting

A problem, new with jet transport aircraft, is that of high-altitude lighting of the aircraft interior. Commercial aircraft will, in the near future invade the altitude levels that have been the exclusive operating area of the military. Two factors are important: First, the pilot will need to be more alert to operate without collision, and second, the internal lighting is reduced in daylight due to the light distribution at high altitude.[11,12]

At high altitude, there is a reduction in the scatter media in the air between any point and the sun.[12] This results in reduced scattered light falling on any point. This scattered light comes to a point along other than straight line path from the source. Thus, it is scattered light that normally enters the windows of the house or office to light up the interior. In the past, scattered light has been the source for daylight illumination of the interior of an airplane cockpit. As operating altitude is increased, there is a reduction is scattered light and the sky becomes darker. The scatter media is now below the aircraft and appears as an illuminated surface, a cloud. This has been called the "milk bowl" effect.[4] The pilot seems to be in a large inverted black bowl over a sea of milk. There is a single point source of light, the sun, and its intensity is increased to perhaps 12,000 foot-candles, probably 25% more than normal at ground level.[1,2] This produces the high-altitude lighting situation as follows: The pilot faces a greater glare source if there is any object in his field of vision, i.e., clouds or other aircraft. The pilot may be subjected to brightness levels of 8,000 foot-lamberts or more. This causes a higher adaptation level. The interior of the cockpit is at a lower illumination level due to absence of scatter light. The resulting light ratio is extremely high. The effect without high-intensity illumination has been called the "tunnel" effect.

The milk bowl effect intrudes on the pilot's ability to see since the eye is subjected to an abnormally high brightness ratio from below. There is no eyebrow or frontal bone to protect the eye from this 5,000-foot-lambert or brighter source. The result is higher adaptation level and reduced visual ability in the cockpit.

To maintain good vision in this high adaptation level, an attempt has been made to bring up the interior light levels to 125 foot-candles on operating surfaces and the main instrument panel. This will bring the maximum brightness ratio in the range of 65 to 1. This ratio is very high for close work. The normal expected ratio will be less.

Styling Considerations

Styling is a major part of lighting design, particularly in the cabin area. This is best described by a term in use among aircraft designers, the "$5,000,000 look." This factor has been in the minds of all design groups but particularly the equipment and interior designers. The entry of a completely new aircraft into the transport field calls for new styling so that the aircraft looks new and different. Departures in styling have rendered much of the lighting of older aircraft unusable. New approaches to lighting are required to maintain the plush appearance of new interior styling. This must improve passenger comfort and safety as well as look new. The fact of larger numbers of people in restricted areas requires better illumination engineering to maintain good visual conditions and to reduce annoyances such as glare sources.

Cockpit lighting requires, where possible, the "$5,000,000 look." There must be a strong effort to make the cockpit pleasant and plush in addition to supplying good lighting. This requires that the lighting be unobtrusive and adequate. To accomplish both features by night and by day and still maintain the "smooth" look is a sizable task.

Electrical and Mechanical Design

Generating System

The change from a 28-volt d-c to a 115-volt a-c generating system as the prime power source is the major reason for the increase in the amount of light supplied in newer aircraft. This permits the use of 115-volt lighting assemblies without excessive weight penalties. The maximum capacity of the 28-volt d-c systems has been limited. The use of additional blocks of power for lighting was frowned upon. Light units using 115-volt a-c power were premium units because of the premium power used. Alternating voltage required the use of inverters with an efficiency of not more than 45% and a weight of $2^1/_2$ pounds per 100 volt-amperes. These factors have limited the lighting engineer in his freedom to exploit new light sources and styles. The advent of the 115-volt a-c system makes 115-volt light units practical and places a penalty on the use of 28-volt lamps, because of transformer weight. The use of larger amounts of light and the use of 115-volt units, particularly fluorescent lamps, are the result of both mechanical and electrical factors.

In selecting a light source for a new aircraft, the field is wide open, even to the use of arc lights. There are three types of light that are presently practical or will be practical in the near future: incandescent, fluorescent, and electroluminescent. The electroluminescent sources have been investigated and found impractical at this time. However, their use can be expected in aircraft in a couple of years. The relative merits of incandescent and fluorescent lights were examined. Fluorescent lights were found more desirable for most interior lighting purposes where extremely small size was not a factor, since efficiency, color, and light pattern were in their favor.

Mechanical Considerations

Mechanical considerations include weight, vibration, size, pressurization, aerodynamic, maintenance, and structural problems.[13] Aerodynamics is a problem of exterior light fixtures and concerns the fixture manufacturer as well as the airframe manufacturer. Today, most external light fixtures are custom-designed and built for each aircraft to secure coordinated design.

Weight and vibration are major factors in all lighting. Vibration is of particular concern since the vibration characteristics of jet aircraft are relatively unknown and since the new lights are to be 115-volt where ever possible. The structural integrity of a filament lamp is a function of voltage and power. An increase in voltage with no change in power or a decrease in power with no change in voltage requires longer and thinner filaments that are more fragile. The 28-volt lamps in common use have not been

as rugged as is desired. Those lamps normally used for aircraft area illumination have had a rated lamp life of 300 hours. This value was considered by the aircraft and lamp manufacturers to be the best compromise of many design factors. Lamp life has been a source of criticism by the aircraft operators; however, no better compromise has been effected.

Lamp Vibration Tests

The body of knowledge on lamp vibration resistance was very limited. A single report was found that proved helpful.[14] A vibration test was instituted to determine the relative merits of various types of lamps. Since the vibration characteristics of new aircraft were unknown, assumptions were made based on theoretical studies and the characteristics of similar aircraft. The environment selected will not be exceeded. Frequencies up to 2,000 cycles per second and accelerations up to 10 *g* can be expected in most areas of the aircraft. In the trailing edges of the wing, 15 *g* may be expected.

Five lamp types were tested over a frequency range of 50 to 1,200 cycles. The definitive tests were conducted at 10 *g*. Evaluation was on a comparative basis since definitive limits were not available. The lamps tested were: *50A* general service; *50A* vibration service; *50A* rough service; *F13T5* fluorescent, and the *AN 3133-311* 28-volt lamp. Since the *311* lamp represented a known experience factor, it was used as the basis of comparison. Lamps were tested to failure or to 60 hours on a sweep-frequency basis at fixed acceleration.

The *50A* vibration service lamp most closely approached the performance of the *311*. The fluorescent lamp failed due to filament breakage. Since this filament is a low-voltage unit, theory would indicate higher vibration resistance. The lamp manufacturers instituted studies to improve the fluorescent lamp filament.

Glass breakage occurred in the fluorescent lamp as the resonant frequency of the tube was reached. This occurred only at 10 *g*. The test program was carried out with a fuse clip used as a damper at the center of the 21-inch tube.

Aviation Fluorescent Lamps

The possibility of a special line of lamps for aviation provided an opportunity to approach one other problem, the bipin base. Lamps, as proposed for use in the aircraft, were in very tight fitting reflectors in all cases. This served two purposes, light control and ground plane. For satisfactory installation of the lamp, the bipin base requires a 90-degree rotation of the tube. In the reflectors involved, there is not room to grasp the tube and rotate it. Some other base form was desired that would lend itself to easier installation in "safety" sockets. The "ring and button" base of the World War II ultraviolet lamp was deemed satisfactory. This provided a lamp that could not be interchanged with more fragile commercial lamps and a lamp adapted to installation in safety sockets. Lamp manufacturers are currently developing such a line of lamps.

The design of a suitable socket for the lamp was relegated to a fixture manufacturer. The socket must provide positive retention of the lamp, visual inspection of that retention, push-in installation, wiping action of the contacts, nonambiguous installed position, screw terminals for external connection, no-tool release and replacement, and minimum size and weight. A socket is under development that appears to meet all of the requirements.

Lamp and Lighting Requirement

The choice of *T-5* lights was based on space considerations and minimum weight. The very limited space in the cabin light cove made the 5/8-inch diameter the obvious choice. The light level must be higher than with previous lamps. Uniform pattern is a styling requirement. Incandescent lamps in a similar cove light fixture have a record of 300-hour life. One air line mentioned replacing 30,000 lamps a year on one model of aircraft for the cove lights alone. The 200 or more lamps in a cove constitute a maintenance problem aside from the cost of the lamps. The 7,500-hour rated life of fluorescent lamps appears very desirable. This is the lamp life for 86% of the lamps still burning.

The high-intensity white light for the instrument panel required light control, light intensity, and placement with space at a premium. The *T-5* fluorescent lamp meets this requirement.

Heat Rejection of Lights

Heat rejection of incandescent lamps became a problem in the cockpit. The design illumination on the instrument panel is 125 foot-candles. An attempt was made to attain this level with incandescent lamps but 500 watts were required and this heat source so close to the pilot in a confined space was intolerable. Also, additional cockpit lighting added to the heat source. Had incandescent lighting been used in the cabin, the air-conditioning load would have been increased under some conditions.

Power Factor and Interference

The side effects of fluorescent lamps require consideration. These include radio interference, power factor, and system stability. Fluorescent-lamp inductive ballasts have low lagging power factor. Capacitive ballasts have leading power factor nearly as low. If leading power factor loads are large, system stability is endangered on the aircraft primary power system. Since the leading power factor load was small, it did not become a problem. The lagging power factor of the dimming-type ballasts is in the range of 30 to 40%. The weight of the capacitors required to correct the power factor made correction less desirable than the low power factor. Further, the load varies as the lamps are dimmed so the degree of correction will vary with the load. One of the reasons advocated for the superiority of fluorescent lights was the reduced power consumption. The power system designers saw no advantage in this because of the relation of load size to system size. Hence no concern was felt over the lagging-power-factor problem.

Radio interference has been a problem in preheat fluorescent circuits of the past. The amount of radio interference in the present dimming circuit was unknown until a system was assembled. Tests in November 1956, showed a clean spectrum above 30 megacycles and a level below the allowable maximum of military specification *MIL-I-6181B* in the frequency range below 30 megacycles. With the probe touching the lamp, no level greater than 30 microvolts was discovered in the lamp and ballast.

Lamp Replacement

Lamp life is important to the aircraft operator. Lamp replacement represents cost and usually represents time lost. If lamp life is short, the initial cost of the lamp is of small consequence compared to the labor cost of replacing it. Incandescent lamps for aircraft are rated for 300 hours when operated at full voltage. Reports from operators indicate that the lamps have about that experience rate. The *311* 50-watt lamp costs about 35 cents. Thus, lamp cost is $1.17 per 1,000 hours. Add to this $3^1/_4$ lamp changes at an average of 10 minutes per nonscheduled lamp change. Labor at $1.50 per hour costs 83 cents. This increases lamp cost to $2.00 per 1,000 hours or more.

The aviation *T-5* fluorescent-lamp line will have a lamp life of about 4,000

hours for 100% of the lamps in operation. With lamp cost at $1.50 and installation cost of 38 cents per lamp (15 minutes at $1.50 per hour), the price of light becomes 47 cents per 1,000 hours, an appreciable saving as compared to incandescent lamp cost.

Considering the nuisance factor of nonscheduled lamp replacements, 100% scheduled lamp replacement becomes both practical and desirable. Very often, if a lamp fails in flight, it is not replaced until the next stop, and thus is a source of customer annoyance.

Color

Illumination color with incandescent lamps is largely fixed by filament temperature. Any color changes must be accomplished by use of filters. With the use of fluorescent lamps, color becomes obtainable in the wide variety of phosphors available. This permits the use of cool colors for space lighting to reduce tendency toward motion sickness and to increase psychological comfort. These cool colors also have the highest lumen efficiency. Cool white is favored for cabin decor at present. Warmer colors are used for the intimate areas and for reading lights. The use of warm or pink colors in lounges, lavatories, and particularly for mirror lights, produces a healthy pink glow on the skin and flatters the makeup and complexion of the women.

One other color is desirable. For night operation, red light is used to maintain low-level night adaptation. With incandescent lamps, this was obtained by red filters or red glass bulbs. The high cost and uncertain results of red glass bulbs make some other approach desirable. Filters are more predictable than bulbs for color but aging of the material tends to produce lighter colors with time. The desired range of red is one with a sharp cutoff at 610 millimicrons. There is not a filter material with the desired sharp cutoff and high light output in the red range. In examining commercially available red fluorescent lamps, the light emitted appears to have sharp cutoff at 610 millimicrons and another at 670 millimicrons. This was visually checked with a spectroscope. The color does not change with dimming and probably does not age during the life of the lamp. If the cutoff is confirmed in further tests, the red fluorescent lamp is very desirable.

Weight and Dimming

Weight and dimming are treated together as they are very closely connected in aircraft use. Dimming was the first problem attacked. Several firms market dimming units for *T-12* lamps. The problem is to obtain comparable dimming on *T-5* lamps, which dim only under extreme protest. The aircraft operating companies were shown a demonstration of cabin light dimming using incandescent lamps and decided that dimmable lights in the cabin are a definite requirement, with no logical reasons necessary. The necessity of dimmable lights in the cockpit is generally understood. To dim the fluorescent lights, a resonant start circuit was first tried. This would dim *T-12* and *T-8* lamps satisfactorily but not *T-5* lamps. It was considered possible to use *T-8* lamps in the cockpit at some sacrifice in efficiency. For the cabin, the weight and size of the *T-8* were impossible. One of the ballast manufacturers was making parallel studies on a dimming ballast for *T-5* lamps and came up with a successful unit. In June 1956, this unit was demonstrated to the airframe manufacturer and the practicality of fluorescent lamps in aircraft was assured. This started a hectic and intense development on lamp dimming that produced the final units.

The lamp-dimming circuit evolved was essentially a rapid-start circuit with variable voltage dimming. This circuit, in commercial use, utilizes continuous filament heating with an inductance ballast. In the *T-5* dimmer, an autotransformer is included because of the high operating voltage of the 13-watt 21-inch tube. A ground plane is required, just as in all rapid-start circuits. In the dimmer circuit, the ground plane to lamp spacing is somewhat critical if flat surfaces are used. With curved reflectors surrounding the tube, the spacing is less critical. The circuit produces dimming ratios above 500 to 1. There are instability points in the operating range that tend to produce banding and ripple under some conditions. The ballast manufacturer found that any flicker could be eliminated. The use of indirect fixtures mitigated the ripple and banding effects.

Dimming controls for a-c lighting circuits are usually of the variable autotransformer (VAT) type. VAT may be continuous or step type. The continuous VAT generally uses a brush or roller moving in contact with the windings. The life and reliability of this brush or roller has been a problem in the past. Dirt, vibration, mechanical damage, and oxidation tended to produce malfunction. This unit usually produces voltage directly proportional to knob rotation.

One manufacturer attacked the reliability problem by application of a step switch and switching reactor. This permits a choice of voltage-knob rotation curves. The reliability of the rotary switch is stated to be an improvement over previous voltage-varying methods. The reliability is a function of the reliability of the switch mechanism. In the larger power-handling capacities, the step VAT is a space- and weight-saving unit. As the power requirement becomes less, this saving disappears.

The number of steps required on a step VAT will vary with the application. For cockpit light control, nine steps were considered adequate. These were set up on a logarithmic basis following the classic theory of logarithmic eye response. This control provides wide dimming steps at the bright end and permits small light variations at the dim end where vision is critical with respect to light level. Step dimming provides the same advantages for incandescent and fluorescent light installations.

Weight is a consideration in fluorescent lamp ballasts as in all other aircraft components. To be considered for aircraft use, the fluorescent lamp system must be competitive with the incandescent system with respect to weight. This is a long stretch from present 60-cycle commercial lamp ballasts. The weight of incandescent systems includes the heavier wire required for any given power load and the transformers required for conversion to 28 volts. Dimming VATs were left out of the weight analysis on the assumption that the weights will be equal for both systems. This penalizes the fluorescent dimming system. The fluorescent system includes ballast, wire, and fixture weight. Ballast weight is the big problem. The first 4-lamp ballast weighed 5¼ pounds. This was reduced by much effort to 27 ounces; and the effort still continues. The point has been reached where the weight for equivalent light output is about equal for the fluorescent and 28-volt incandescent systems.

Power Consumption

Another consideration is power consumption. Where electric power is in limited supply, the increased efficiency of fluorescent lights in lumens per watt is an advantage. In the case of the transport aircraft, there are other more important considerations. The reduced power consumption was not considered a significant advantage. The relative power consumption advantage of fluorescent lamps is about 1.5 over incandescent lamps, considering all the accessories. In aircraft with a limited power supply, this can be an advantage.

Fixtures and Light Control

The aircraft cockpit required control of light to a greater degree than ever attempted in previous illumination studies. This is because of the vision requirements at night and the high brightness ratios by day, both of which place a strain on a flight crew already burdened by a tremendous responsibility. Fluorescent lamps do not lend themselves to as rigid control as do incandescent lamps. Incandescent lamps present a near point source of light where the fluorescent lamp is widely distributed and diffuse. This diffuse source is desirable in some situations.

Lighting for daylight operations is based on the high-altitude requirement mentioned previously. The desired level is about 125 foot-candles on the main instrument panel. The problem is similar to that of the sign industry, lighting a plane surface from one edge and obtaining uniform brightness. The problem is further complicated by the depth of instrument dials behind the plane of the panel and by the instrument bezels that overhang the dials. This introduces a "tunnel" and the shadow tends to obscure the upper portion of the dial. It is desirable to light this tunnel and to break up any sharp shadows on the instrument panels. The use of fluorescent lamps has effected this very well. The continuous line source of the tube leaves no obvious shadow on the dial and no sharp shadow below projecting objects on the panel.

In the vertical plane, it is desirable to get even illumination over the height of the panel. The fluorescent source is distributed and diffuse but the use of cylindrical parabolic reflectors enables a sufficient degree of concentration to obtain good control. The reflectors in the cockpit were of very short focal length for reasons of space. The cosine distribution of any element of the lamp surface makes the summation appear to emanate from the axis of the lamp. The reflectors were focused on the most distant row of instruments. This partially compensated for the square law effect. With a distance ratio in excess of 4 to 1, the light ratio was kept below 3 to 1. Lamp length became a problem since the glare shield is a curved surface. The 21-inch length of the 13-watt lamp was too long and had to be replaced by a greater number of 12-inch 8-watt lamps. Weight had to be sacrificed for installation.

Fabrication of the complex reflector shapes caused an investigation into Fiberglas laminate techniques. Since the axis of each tube was not parallel to any other tube axis or to any surface, the laminating technique solved the form problem. The inner surface was lined with aluminum foil, laminated to the assembly. The finished part is light in weight and has good light-control characteristics.

Flight Engineer's Panel Lighting

The daylight lighting on the flight engineer's panel presented a problem of control in that the pilot and copilot also wish to look at this panel. This is a very large panel, vertically and horizontally. Any light source used to illuminate the panel must be in such a position or of such a nature that it will not glare in the eyes of the pilot or copilot. On such a large panel, a large amount of light is required. The use of sealed-beam incandescent lamps presented glare source and installation problems. A fluorescent lamp installation reduced the installation problem caused by the small diameter of the lamps.

To secure a degree of light control, the cosine transmission of a diffuse, translucent medium was utilized. Light through such a material has a principal beam normal to the surface. A group of fluorescent lamps were installed with axes fore and aft, each behind a diffuse surface. The surfaces were oriented to place the principal beam on the bottom of the panel, and were joined with an opaque material to form a sawtooth form. The result was a low brightness source, oriented away from the line of vision of the crew members. Because of the length of the lamps, the shadow of the flight engineer is partially nullified, and sharp shadows on the panel are broken up. The panel light intensity is sufficiently high to reduce adaptation time when looking from outside high-brightness source to the panel. The complete light installation is fabricated from Fiberglas using special lay-up techniques. The assembly is equipped with a rack-and-panel connector and becomes a "one-man-installation" item.

Night Lighting

Night lighting of all instrument panels is accomplished by integrally lighted instruments, with low-level background fill light to reduce vertigo. Name plates and instructions are of the back-lighted plastic-plate type, commonly called edge-lighted panels. In both cases, light is introduced into a transparent material. The surface is opaque except for the intelligence to be transmitted. The intelligence lights up against a dark background, producing maximum contrast. Red light is used for maximum night adaptation. Very careful placement of all lighted materials is required to minimize windshield and window reflections. Those items of primary importance should be located to eliminate reflections. Less important items are located to minimize reflections and are turned off when not in use.

The design and manufacture of plastic plates is a subject in itself. The use of electroluminescent sources was investigated for this application. The present state of the art does not warrant their use.

In the interest of space economy and reduced installation cost, integrally mounted light fixtures and integral connections were used in the plastic plates. The color selected for the plastic plates is a compromise of many factors. In the past, plates have had white letters on a black background for maximum contrast in daylight. The black plates tend to reduce the brightness level of the cockpit as compared to a lighter color. A dark gray was selected as background color. This produced satisfactory readability under the higher lighting levels available. It also matched the styling and esthetics of the cockpit.

Miscellaneous Cockpit Lights

The miscellaneous lighting of the cockpit area includes such items as floor lights, work lights, map lights, and chart board lights. Each of these requires a light for a special purpose. Where possible, 115-volt lamps are used to eliminate transformer weight. The cockpit floor lights use a special incandescent 115-volt lamp *10C7/1DC*. This is a sewing machine lamp with a ruggedized filament. A fluorescent lamp is used for the navigator's work light. Where these lamps are not usable, 28-volt lamps are used.

Lighting Levels

Cabin Lighting Levels

The assignment of definite lighting levels for purposes other than reading lights has been avoided in previous aircraft, possibly because of inability to meet any adequate light level. The lighting levels attained on modern aircraft are the result of many studies by various groups. The light engineering of the cabin was contracted to a transportation lighting firm. This firm consulted with the design groups and industrial designers as necessary. The lighting levels in the cabin are based on the experience of this firm. Later measurements

will confirm the existence of adequate lighting.

The requirements for dimming of the main cabin lights was an obstacle until successful fluorescent light dimming was demonstrated. The reasons for this requirement were developed by a process of rationalization after the desire for cabin light dimming existed. Dimming was rationalized as less of a shock on the passengers due to turning the lights on and off abruptly. The optimum lighting levels in the cabin are presumed to be higher because of the high-altitude operation, for reasons noted previously.

Reading Light Levels

Reading lights are always a major consideration from the standpoint of passenger comfort. Since many passengers pursue paper work throughout the duration of the flight, light levels should be adequate for this activity. In the past, a level of 25 foot-candles at the surface of the paper has been a design criteria. Pattern was largely a matter of what could be done with inside frosted lamps, reflectors, and lenses. The usual incandescent fixture has a small spot of high intensity in the center and tapers off to the 6- or 8-foot-candle level 8 or 10 inches from the center. Fluorescent fixtures for reading lights have been investigated extensively. One form is being used in a production aircraft. Much remains to be discovered before fluorescent light fixtures reach an ideal state of development. Large-area uniform light patterns may result from continued development of fluorescent-equipped fixtures.

Analysis of Light Levels

Many light levels are not given a rigid analysis. A light fixture is installed and one or more people visually judge the adequacy of the light. Usually it does look like enough light and the fixture is used without further analysis. Other areas are of sufficient importance that they are given an analysis of sorts. This is the case with wheel well lights and baggage compartment lights. These are areas where adequate lighting did not exist in the past. Referring to some standard work on lighting levels, a value is selected on the basis of similarity of function.[8] An attempt is then made to assure this light level by use of proper bulb size and reflectors.

Some light levels are the subject of regulation. Exterior navigation lights are specified by Civil Air Regulations as to pattern, color, and light output.[15,16] The navigation light fixtures usually are specified by the fixture manufacturer as meeting the regulations.

Light levels in some areas are the subject of very extensive study prior to their establishment. The high-altitude lighting of the cockpit was the subject of such a study. There is little information on the lighting of cockpits at high altitude. The existence of the problem was not recognized prior to a study on cockpit illumination as a function of color. Sky brightness was examined in connection with the color study. Literature pointed to the reduction in sky brightness with increased altitude and pointed out the effect of the visual field.[1–3,17] Some work has been done on attempting to correlate sky brightness with cockpit illumination. This work has not been completed since testing at high altitude is required. Sufficient analysis was accomplished to enable the selection of lighting values by intuitive "guestimation." The results will be best evaluated when the aircraft is taken to 40,000 feet and pilots report back.

Landing-Taxi Lights

Landing and taxi lights are designed around one or more available lamps. The lamps in present use are of the sealed-beam *PAR* type. The landing–taxi light system serves to light the path of the landing aircraft and to light the area around the taxiing aircraft. The landing function requires a high-intensity narrow beam, straight ahead. Lamps in current use have an intensity of 550,000 beam candlepower or more in the 8-inch-diameter lamps. Taxiing requires a high-intensity beam ahead with reduced light levels to the sides. The forward beams for landing and taxiing require lamps with the surface mounted normal to the line of travel. To make this lamp aerodynamically acceptable, it is mounted in a retractable fixture and extended only at reduced speed or as needed. A 2-filament lamp is available with a 600-watt narrow landing beam and a 400-watt wide taxi beam. This reduces the requirement for extra installations and extra weight for the taxi lamp. Additional *PAR 46* lamps are installed in surfaces that permit the lamps to shine out to the side without protruding into the airstream. For landing operation, all the lights are turned on. This provides a high-light-level landing beam. For taxiing, the high-intensity landing beam is turned off, thus conserving the life of this high-temperature filament. This provides higher intensities for both landing and taxiing than were previously available.

Design Carry-over

A factor of considerable importance in any design is carry-over. This enables much of the design to be repetitive and conserves manpower for items requiring new emphasis. Decisions on lighting were materially influenced by this consideration. The problem is more difficult since a change in primary power is involved. The size difference between 28-volt and 115-volt lamps makes every 115-volt fixture a developmental item. If scarce engineering manpower is to be conserved, a choice must be made between the use of an existing 28-volt fixture or the elimination of transformers by development of a suitable 115-volt fixture. Some locations, plastic plate lamps for instance do not permit the use of 115-volt lamps because of space limitation. Factors such as these dictate the use of 28-volt lamps even though the primary power system is 115 volts.

Conclusions

The use of fluorescent lamps in transport aircraft presents advantages in lamp life, color, light pattern, and power consumption. The fluorescent lamps are the equivalent of 28-volt incandescent lamps in system weight. Where very-small-size light sources are desired, the 28-volt incandescent lamps are available and fluorescent sources are unusable. Developments over the past 2 years have made available fluorescent lights that are vibration-resistant and ballasts that are lightweight and enable dimming of the fluorescent lights. The existence of 115-volt primary power makes the use of fluorescent lights practical.

The use of adequate amounts of light in all locations is now possible with primary power systems of sufficient capacity. Convenient dimming of all types of lights is possible with alternating voltage as primary supply. Reduced scatter light at high altitudes reduces the light flux into the cockpit. The availability of larger amounts of dimmable light with reduced heat rejection makes more palatable the higher artificial light flux required for high-altitude operation.

Additional research and development in the field of air transport lighting will further increase air-crew efficiency and passenger comfort.

References

1. Brightness and Polarization of the Daylight Sky, Tousey, Hulburt. *Journal*, Optical Society of America, New York, N. Y., vol. 37, Feb. 1947, p. 78.

2. Brightness and Polarization of Daylight Sky to 35,000 Feet. Packer, Locke. *Ibid.*, vol. 41, 1951, p. 473.

3. Brightness of the Atmosphere, N. L. Barr, et al. *Research Report*, Project NM 001 056.07.01, Bureau of Aeronautics Project TED PTR AC223, Washington, D. C., vol. 11, Mar. 1953, p. 191.

4. Brightness of the Atmosphere, N. L. Barr, et al. *Research Report*, Project NM 001 056.07.03, *Ibid.*, Oct., p. 1011.

5. Visibility of Airport Runways, N. L. Barr, et al. *Research Report*, Project NM 001 056.07.53, *Ibid.*, vol. 12, Nov. 1954, p. 435.

6. Report to the Indianapolis Air Collision Symposium, P. J. Sutro. Joint Civil Aeronautics Authority–Illuminating Engineering Society Symposium, Indianapolis, Ind., Nov. 8, 1955.

7. Visual Effect of Nonuniform Surrounds, M. Moon, A. Spencer. *Journal*, Optical Society of America, vol. 35, Mar. 1945, p. 233.

8. Handbook of the IES. Illuminating Engineering Society, New York, N. Y., second edition, 1954.

9. Visibility of Distant Objects, Silvert Q. Duntley. *Journal*, Optical Society of America, vol. 38, Mar. 1948, p. 326.

10. Dark Adaptation Following Light Adaptation to Red and White Light, Hecht, Hsia. *Ibid.*, vol. 35, Apr. 1945, p. 261.

11. The Physical Environment of the Flyer, Heinz Haber. U. S. Air Force School of Aviation Medicine, San Antonio, Tex., 1954.

12. Visibility of Cockpit Instruments, N. L. Barr. *Technical Data Digest*, Dayton, Ohio, vol. 15, no. 10, p. 26.

13. Better Lighting for Tomorrow's Planes, P. E. Massie. *Aviation Age*, New York, N. Y., vol. 26, Apr. 1956, p. 46.

14. Factors Affecting the Performance of Indicator Lamps, H. M. Ferree, J. D. Lash. *Illuminating Engineering*, New York, N. Y., Sept. 1951 (abstract).

15. Study to Determine Optimum Exterior Lighting, M. K. Laufer. *WADC Technical Reference 55-309*, Wright Air Development Center, Dayton, Ohio, Oct. 1955.

16. Civil Air Regulation 4b.636–8. Civil Aeronautics Authority.

17. Lighting Design (book), M. Moon, A. Spencer. 1948.

The Advantages of a Pneumatic Electric Power System for High Mach Number Aircraft

MORTON A. SLAVIN
NONMEMBER AIEE

AS THE SPEED of aircraft increases, the capability of the accessory power equipment to meet the added requirements imposed by such vehicles must also increase. The important need is for these capabilities to occur in time to be used on the aircraft.

One of the most difficult problems created by the increasing speed of aircraft is the accompanying increase in ambient temperature. The choice of one type of accessory system over another could be determined by the amount of cooling each would require to meet these high-temperature applications. In most cases the greater the cooling requirement, the greater is the penalty in aircraft performance.

A pneumatically driven electric power system with the capability of operating today in an environment created by Mach 3 flight is presented in this paper. The system is based on the use of a turbonator recently developed by the General Electric Company and discussed in a separate paper.[1] The turbonator, an integral turbine wheel and generator, requires no external cooling for the two major heat-dissipating components, the drive and the generator. In addition, this system can be designed so that the penalty due to power extraction can be made competitive with engine shaft powered accessory systems. This pneumatically driven electric system is therefore capable of reducing airplane cooling penalties at no expense in power extraction penalties.

Paper 58-910, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 10, 1958; made available for printing April 18, 1958.

Morton A. Slavin is with the General Electric Company, West Lynn, Mass.

The author wishes to thank Dr. L. Schindel, Dr. E. Rubin, and Dr. E. Covert, associated with the Analyses and Research Division of the Aeronautical Department of Massachusetts Institute of Technology, for their valuable assistance in preparing this paper.

Discussion

All accessory systems impose a penalty on the performance of an aircraft, since the general method of evaluating accessory systems is by comparing their penalty to aircraft performance with the base line established by an aircraft with no accessories.

Some of the major factors involved in an accessory system evaluation are weight, cooling, volume, and power extraction. However, other factors such as reliability, maintainability, location, etc., must also be considered. It is the summation of all these effects that dictates the type of accessory system best suited for a particular application.

This paper will confine itself to an analysis of the penalties associated with a pneumatically driven electric system for a high-speed aircraft and a comparison of these penalties to those of a shaft-driven electric system. These systems are designed with components, available today. The major component of the pneumatic system is the turbonator, which is a single package containing a 24,000-rpm turbine wheel and generator on a common shaft. The shaft-driven system is based on the ball pump hydraulic constant-speed drive. Both the turbonator and the hydraulic constant-speed drive have been built and tested by the Aircraft Accessory Turbine Department of the General Electric Company.

The effects of engine power extraction, cooling, weight, and volume of this system on aircraft performance are determined for an assumed application. These effects are compared to those obtained by a shaft driven hydraulic constant-speed drive and and its generator.

This study assumes a Mach 3 airplane with 90-kva (kilovolt-ampere) generator requirement. The penalties for two types of aircraft are analyzed: an interceptor type with a single engine and a single 90-kva generator, and a bomber type with six engines and four 90-kva generators. The engine has been assumed to be General Electric's *X279* whose performance was used throughout the study. The changes in engine performance due to accessory power extraction are small, even for a 90-kva load and, to assure the necessary accuracy, the engine characteristics were obtained on a digital computer with the use of the customer available engine performance computer cards.

Since information concerning the engine performance is classified, the results are presented in the form of a percentage change in aircraft performance due to the accessory power system characteristics.

The turbonator design for this theoretical application is similar to that used

for the demonstrator unit now on test. The individual components have the same temperature limitations as those already tested. The hydraulic constant-speed drive and generator are also similar to those in use today. The size and weight of each have been extrapolated from existing data.

The major portion of the accessory power system penalty to aircraft performance is realized during that steady-state portion of the mission which has the greatest duration. This study has assumed that a Mach 3 dash will be the major portion of the mission.

The basic feature of the turbonator which gives it the high-temperature capability is the ducting of the turbine exhaust air for the pupose of cooling the generator. Maximum capability of the turbine temperature drop is obtained by the use of a partial arc turbine nozzle design which allows for maximum turbine efficiency at the high-speed design point. These relatively cool turbine exhaust temperatures are obtained only when operating at the design electric load. When there is no electric load, there is theoretically no temperature drop across the turbine. It is this operating condition which fixes the inlet air temperature that this particular generator design will allow. The maximum inlet air temperature that can be tolerated by today's machine is that obtained with ram air at Mach 3 flight at altitude. The flight speed can be increased above Mach 3 if an electric load is always available either in the form of a real or a "dummy" load or if an intercooler is used for cooling the turbine inlet air with the turbine exhaust air. Obviously, the flight Mach number can be increased if a higher temperature generator insulation is developed.

Therefore a bleed-to-ram air-switching device is required so that when the speed is reached where the bleed air temperature is too high to cool the generator, the inlet air will be switched to ram. Such a switching device has only to sense ram pressure, temperature, and pressure ratio to perform this operation adequately.

The penalty of cooling the generator voltage controls and the drive-speed-sensing controls is not considered. Control systems using today's electric components will require cooling for these ambients. Since all electric systems will require essentially the same type of control, the cooling penalty will be the same for all systems. In any case the losses of the control system are relatively small and thus the cooling penalty for any system is small.

Method

The effect of power extraction, weight volume, and cooling requirement of each accessory power system on the performance of the aircraft is obtained by analyzing each factor separately. The total penalty to the performance of the aircraft will be the sum of each one.

The method by which each penalty was calculated is presented in the Appendix. The assumption required and the equations used are worked out for each characteristic of the accessory system for a bomber-type airplane and for an interceptor-type airplane.

In essence, it was assumed that a fixed-design bomber-type airplane would fly the same mission with and without each of the two electric systems. The initial gross weight of the aircraft is fixed and it flies its mission at a constant Mach number and a constant lift-drag ratio. The bomber is assumed to have six engines, a gross weight of 400,000 lb (pounds) and an empty weight of 200,000 lb. The electrical requirement is for four 90-kva generators, each of which operates at 50% load during the Mach 3 dash. The penalty of each electric system is the decrease in the range caused by the weight, size, cooling, and power extraction requirements of the accessory system.

The power extraction penalty is expressed in the form of per-cent decrease in range due to the increase in specific fuel consumption which results when power is extracted for the accessory system. In all cases the specific fuel consumption is based on a constant value of thrust in order that the aircraft may maintain a constant Mach number. Thus a different engine throttle setting is required for each case of power extraction in order that the net thrust remains equal to the value obtained at zero power extraction.

Table I. Per-Cent Decrease in Range for Bomber-Type Plane

Source of Penalty	Electric System No.* 1	2	3
Power extraction	0.85	0.24	0.31
Equipment weight	0.81	0.77	0.99
Equipment volume	0.40		
Equipment cooling per hour			0.12
Totals	2.06	1.01	1.42†

* Note:
System 1, Ram-air-driven turbonator exhausting to ambient.
System 2, Ram-air-driven turbonator exhausting into engine secondary air.
System 3, Engine-shaft-driven hydraulic constant-speed drive and generator.

† Cooling portion of this total is on a per-hour basis.

The weight penalty of an accessory system is determined by decreasing the aircraft fuel weight an amount equal to the weight of the accessory system. The penalty is therefore the decrease in range resulting from the loss in fuel.

The size or volume penalty depends greatly on each particular application. For this study it was assumed that the volume of the accessory system displaces an equal volume of fuel. As in the case of the weight penalty, the decrease in range results from the loss in fuel. Since the volume of the turbonator, and hydraulic constant-speed drive and generator are essentially equal, the volume penalty arises only for the pneumatic ducting.

The cooling penalty applies only to the hydraulic constant-speed drive system since the turbonator requires no external cooling. It may be assumed that a Mach 3 aircraft is temperature-limited, thus requiring air conditioning for some of its components. If the hydraulic constant-speed drive and generator unit is cooled by means of the engine fuel, then some other accessory requires air conditioning. If a turbonator is used, then extra cooling capacity is available in the fuel for the components previously requiring air conditioning. It is therefore necessary to penalize the hydraulic constant-speed drive and generator unit an amount equivalent to its cooling requirements.

One measure of the cooling requirements is the quantity of water necessary to carry away the losses. This amount of water results in a range penalty by displacing an equal volume of fuel.

The interceptor-type airplane penalty due to the accessory power systems is presented in the form of a decrease in the rate of climb ability rather than a decrease in range. An interceptor mission is usually more critical to net thrust or speed. The rate of climb factor which is a measure of these factors is simply the

Table II. Per-Cent Decrease in R/C Factor for Interceptor-Type Plane

Source of Penalty	Electric System No.* 1	2	3
Power extraction	1.04	0.29	0.34
Equipment weight	0.63	0.58	0.75
Equipment volume	1.31		
Equipment cooling			0.80
(Flight time)			(1 hr) (3 hr) (5 hr)
Totals	2.98	0.87	1.89 2.87 3.66

* See Note of Table I.

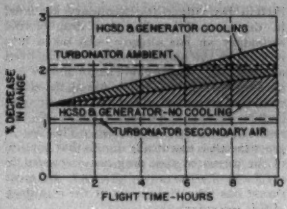


Fig. 1. Per-cent decrease in range versus flight time at Mach 3, for a bomber-type airplane, resulting from various types of electric systems

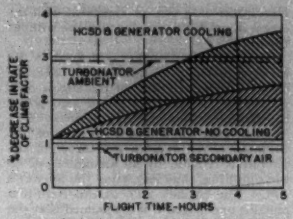


Fig. 2. Per-cent decrease in rate of climb ability versus flight time at Mach 3, for an interceptor-type airplane, resulting from various types of electric systems

excess thrust to gross weight ratio for each characteristic of the electric systems. It is assumed that this interceptor has a single engine and weighs 40,000 lb with an empty weight of 20,000 lb. The electrical requirement is for one 90-kva generator operating at 50% of rating during the Mach 3 dash.

The power extraction penalty for an interceptor-type airplane is expressed as the decrease in its rate of climb ability due to the decrease in thrust resulting from power extraction. The decrease in rate of climb ability due to accessory equipment weight results from the increase in gross weight of the aircraft an amount equal to the weight of the accessory system.

The volume penalty again depends greatly on the particular application. An interceptor-type aircraft is a high-density design. It is assumed for the study that the volume displaces an equal volume of fuel. However, to maintain equivalent mission capability, the displaced fuel is carried in external fuel tanks. The decrease in the rate of climb ability is therefore due to the decrease in thrust resulting from the drag of the external fuel tanks.

The cooling penalty for an interceptor-type aircraft is determined in a manner similar to that for a bomber. However, the decrease in the rate of climb ability is based on both the weight of the water, which increases the gross weight, and the drag of the external fuel tanks required to carry the water, which decreases the net thrust.

For each type of aircraft, two types of turbonator electric systems are considered, each having distinct advantages and each causing different aircraft performance penalties. The first system is so installed that the turbine exhaust is fed directly into the ambient slip stream. The second system differs in that the turbine exhaust air is ducted directly into the engine secondary air stream. The latter system is most feasible if the turbonator is engine-mounted, which would be directly in the secondary air stream. The main advantage gained by this type of installation is in the recovery of ram air energy in the form of thrust, thus reducing the drag penalty on the aircraft.

For simplication, it is assumed when calculating duct weigh⁺ and volume that an engine-mounted tu onator exhausts into the engine secondary air stream and a remotely located turbonator exhausts into the ambient slip stream. The engine mounted system therefore has no duct volume penalty since this volume is part of that required for the secondary air flow.

Summary of Penalties

The various range penalties for each system may be added to obtain a total penalty. These values are shown in Tables I and II and are plotted in Figs. 1 and 2.

Conclusions

1. Generating secondary power abroad high-performance aircraft has a small effect on aircraft performance. Since the effect on aircraft performance is the small difference between some very large numbers, a slight error in these large numbers can cause penalty differences as great as the penalties themselves. To obtain the relative change in aircraft performance for various accessory systems, computer-type evaluations are required.

2. For a reasonable bomber mission of over 4 hours in duration, the hydraulic constant speed drive system and the turbonator system exhausting to ambient are essentially a standoff with respect to range penalty. The two systems are also a standoff for an interceptor-type plane.

3. Greater consideration in the selection of a secondary power system should be given to a complete evaluation of such factors as maintenance, location, reliability, etc.

4. The choice of the electric system will also depend upon what other accessory systems are required. A similar analysis such as this should be performed for the entire accessory system.

5. For missions of long duration, the turbonator system shows an advantage as the cooling penalty of the hydraulic constant speed drive and generator becomes quite large.

6. A turbonator system which exhausts into the engine secondary air stream can have significantly lower penalties to the aircraft performance than either of the other two systems.

Appendix

The derivations of the equations used to determine the various penalties in this report follows. Although the equations are essentially based on Brequet's range equation, the derivations will clearly show the assumptions made for this study.

Symbols

W = instantaneous weight of airplane, lb
f' = fuel rate, lb/hr (pounds per hour)
t = time, hr
SFC = specific fuel consumption, lb/hr/lb of thrust
F = net thrust, lb
D = drag, lb
L = lift, lb
V = velocity, miles per hour
Wf = gross weight of airplane, lb
We = empty weight of airplane, lb
Wa = accessory system weight, lb
Wa' = weight of fuel displaced by the volume of the accessory system
Ww = weight of cooling water, lb
Q = rate of cooling water, lb/hr
R/c = rate of climb
a (subscript) = value with accessory system

Bomber-Type Airplane

It was decided to show the various penalties to the performance of a bomber-type aircraft in the form of per-cent change in range. The equation for range is derived as follows:

The change in weight of the airplane with respect to time is equal to the rate of fuel consumption.

$$\frac{dW}{dt} = -f' = -(SFC)F$$

For steady state flight, the following is true:

thrust = drag

$$F = D$$

lift = weight

$$L = W$$

$$\frac{dW}{dt} = -(SFC)D = -(SFC)\frac{D}{L}W$$

$$\frac{dW}{W} = -(SFC)\frac{D}{L}dt = -(SFC)\frac{D}{L}\frac{dt}{dR}dR$$

Assuming that the mission is flown at constant velocity

$$\frac{dR}{dt} = V$$

$$dR = -\frac{V}{SFC}\frac{L}{D}\frac{dW}{W}$$

It is reasonable to assume that a mission can be flown at a constant lift-drag ratio if a constant velocity is maintained. This can be done by increasing altitude as the aircraft weight decreases with fuel consumption. The *SFC* can be assumed constant if an average value is used based on the varying altitude of the mission. For the study the values at 70,000 feet were assumed as representative of the average *SFC*. In general, the change in *SFC* at a constant speed and constant engine setting does not vary significantly with altitude. It is realized that these assumptions are not precise if absolute range calculations are required. However, for purposes of evaluating the difference in range caused by one accessory system over another, these assumptions are permissible.

$$R = -\frac{V}{SFC}\frac{L}{D}\int_{Wf}^{We}\frac{dW}{W} = \frac{V}{SFC}\frac{L}{D}\ln\frac{Wf}{We}$$

where *Wf* is the gross weight at start of mission and *We* is empty weight of airplane theoretically occurring at the end of the mission. The difference between these values is the weight of fuel.

Change in Range Due to Accessory Power Extraction

The change in range due to power extraction results from the change in *SFC*. This assumes that the airplane flies the identical mission with and without power extraction, with the same velocity, lift-drag ratio, initial gross weight, and fuel weight.

$$\frac{\Delta R}{R} = \frac{R - Ra}{R} = 1 - \frac{(SFC)}{(SFC)_a}$$

where the subscript *a* refers to value with power extraction.

The value of *SFC* to be used with power extraction must be such that the net thrust is equal to the value obtained with zero power extraction in order that airplane velocity may be kept constant. To keep thrust constant, the throttle setting must be changed when accessory power is extracted. This assumption is valid only if the throttle setting at zero power extraction is less than maximum. For this study, therefore, this throttle setting is assumed to be at minimum reheat.

The change in range due to power extraction can therefore be expressed in terms of fuel flow only:

$$\frac{\Delta R}{R} = 1 - \frac{f'}{f'a}$$

Turbonator Using Ram Air Exhausting to Ambient

The use of ram air does not directly affect engine performances as such, but it does increase the ram drag of the airplane. This increase in drag can be expressed as a decrease in net thrust for the airplane without the drag.

The *SFC* of an engine is defined as the ratio of fuel flow to net thrust. The aircraft's *SFC* equals that of the engine *SFC* when there is zero ram air or zero power extraction for the accessory system.

When ram air is taken aboard for the turbonator, the over-all net thrust of the aircraft is reduced by an amount equal to the rain drag, thus increasing the aircraft's *SFC*. The engine's *SFC* at this condition remains unchanged if it is assumed that the aircraft speed and engine speed are unchanged. However, this assumption is a poor one since the aircraft will slow down with the loss in thrust. Therefore the engine thrust must be increased by advancing the throttle setting until it is equal to the original net thrust at zero power extraction plus the drag of the turbonator ram air.

The aircraft's *SFC* is now equal to the engine's higher fuel flow at this new throttle setting, divided by the aircraft's net thrust, which is still equal to that value obtained at zero power extraction.

The per-cent change in range due to taking ram air aboard is equal to one minus the ratio of the *SFC* of the airplane with zero ram air to the *SFC* of the airplane with ram air:

$$\frac{\Delta R}{R} = 1 - \frac{SFC}{(SFC)_a}$$

The quantity of ram air required to carry 50% of the rated load at the Mach 3 flight condition is 0.52 lb/second based on a generator efficiency of 82%. The ram recovery assumed at Mach 3 is 72%. The resulting decrease in range calculates as 0.85%.

Turbonator Using Ram Air Exhausting into the Engine Secondary Air

The effect of exhausting turbonator air into the engine's secondary air stream is to recover some of the ram air energy in the form of thrust and thus reduce the ram drag penalty associated with powering the turbonator.

The ram drag to be charged against aircraft performance for this case is the ram drag of the turbonator air flow minus the thrust obtained by expanding the exhaust air through the engine jet nozzle. This expansion occurs at the assumed secondary air pressure to ambient pressure ratio of 8 to 1. This ratio is usually lower than the ram pressure ratio due to losses in the passage between the engine and its nacelle and also due to dump ports which maintain a low pressure for mechanical limitation of the nacelle and variable area engine nozzle. This back pressure of 8 to 1 requires a greater air flow for the turbonator than the one exhausting to ambient since the pressure ratio available across the turbine has decreased. Although this increase in ram air increases the drag, the net ram drag resulting after subtracting the recovered thrust is less than that experienced with the turbonator exhausting to ambient.

The method used for calculating these values is similar to that used by General Electric jet engine departments. The gross thrust quoted for the engine assumes a certain percentage for secondary air flow, in this case 7%. The ram drag of the secondary air has already been taken into account in the quoted thrusts. It has been assumed that the engine ejector nozzle can pass up to 20% secondary air at the flight condition assumed. Any increase in secondary air over the design value of 7% will result in a loss in gross thrust. This loss in gross thrust is essentially due to that portion of the increased ram drag not recovered in the form of thrust. In all cases, the gross thrust has taken into account the ram drag of this secondary air.

The gross thrust can be obtained at any secondary air flow by means of an equivalent thrust coefficient. This thrust coefficient defined below decreases 2% when the secondary air increases from the design value of 7% to 20%.

$$\text{Equivalent thrust coefficient} = C^* = \frac{\text{actual gross thrust}}{\text{theoretical gross thrust}} = \frac{F}{F'}$$

The actual gross thrust is, therefore, that obtained by both the engine's primary air W_B and secondary air W_S minus the ram drag D of the secondary air.

$$C^* = \frac{F(W_B + W_S) - D(W_S)}{F'}$$

The turbonator air flow for this system has increased to 0.72 lb/second. The increase in secondary air flow by this amount results in a decrease in engine gross thrust. This reduction in thrust is the ram drag penalty charged to the turbonator. The aircraft net thrust is obtained by subtracting only the ram drag of the engine primary air flow.

As mentioned before, this operation results in a loss in thrust as well as an increase in *SFC*. Therefore the engine throttle setting must be advanced until the resulting gross thrust or, more accurately, the resulting net thrust is equal to that obtained with zero turbonator power extraction.

However, the increased fuel flow at the new throttle condition results in a higher *SFC*. This higher *SFC* causes a decrease in range as determined by the same equation of one minus the ratio of *SFC*. The resulting per-cent decrease in range calculates as 0.24%.

Hydraulic Constant-Speed Drive and Generator Using Shaft Power Extraction

Powering accessory systems directly from the engine by means of either a mechanical shaft or compressor discharge air results directly in a change in the performance of the engine. As in the previous analysis, the thrust as well as the *SFC* changes as shaft power is extracted.

It was assumed that the amount of shaft power required for this study to carry 50% electric load is 67 hp (horsepower). This value is based on a drive efficiency of 85%, generator efficiency of 82%, and engine gear box efficiency of 97%.

Various throttle settings were run on the International Business Machines Corporation computer with 67-hp extraction. The *SFC* was obtained at the throttle setting which resulted in a net thrust equal to that obtained with zero shaft extraction. The

resulting SFC caused a per-cent decrease in range of 0.31%.

Change in Range Due to Accessory Equipment Weight

The change in range due to the addition of equipment weight to the airplane results from its displacing an equal weight of fuel. Again it assumes that the airplane flies identical missions with and without the additional equipment weight with respect to speed, lift-drag ration, SFC, and gross weight.

$$\frac{\Delta R}{R} = -\frac{\ln \frac{Wf}{We+Wa}}{\ln \frac{Wf}{We}}$$

where Wa is the accessory equipment weight.

The estimated weight of a 90-kva hydraulic constant-speed drive and generator is 300 lb. This weight does not include the weight of oil, heat exchanger, and engine accessory gear box. The weight of a 90-kva turbonator is 225 lb.

The weight of the duct must also be included in the total weight of the turbonator system. It is assumed that each engine-mounted turbonator requires a ram air duct 5½ inches in diameter and 5 feet long plus a bleed duct 5 inches in diameter and 1 foot long. For a fuselage-mounted turbonator, it is assumed that the ram air duct increases to 30 feet for a bomber-type aircraft with no change in the bleed duct which is teed into the ram duct. The duct weight of the engine-mounted system is approximately 9 lb including fixtures and the duct weight of the remotely located turbonator is 45 lb.

For simplifying the presentation of data, it is assumed that the turbonator system which exhausts into the engine's secondary air is engine-mounted and the turbonator system which exhausts to ambient is remotely located. By inserting the values of the equipment weight Wa in the foregoing equation the per-cent decrease in range can be obtained.

$$\frac{\Delta R}{R} = 1 - \frac{\ln \frac{Wf}{We+Wa}}{\ln \frac{Wf}{We}}$$

The resulting per-cent decrease in range for the various cases are as follows:

Bomber with Four Turbonators	Decrease in Range, Per Cent
Engine-mounted	0.77
Remotely located	0.81
Hydraulic constant-speed drive and generator	0.99

Change in Range Due to Accessory Equipment Volume

It is extremely difficult to evaluate the range penalty due to volume differences between various accessory systems unless a specific application is considered.

The volume of the hydraulic constant speed drive and its generator is approximately the same as that of the turbonator. Therefore there is no relative penalty between these major components.

However, the turbonator system requires ducts which are additional equipment contributing toward a range penalty. For this study, the duct volume is considered as displacing an equal volume of fuel. The range penalty is found by a similar method as was the equipment weight penalty. For this case the gross weight of the aircraft is decreased by an amount equal to the fuel displaced since the increase in the aircraft empty weight due to the duct weight is negligible.

The displaced fuel is therefore accounted for by a decrease in the gross weight of the airplane.

$$\frac{\Delta R}{R} = -\frac{\ln \frac{Wf-Wa'}{We}}{\ln \frac{Wf}{We}}$$

where Wa' is the weight of the fuel displaced by the volume.

If the turbonator is engine-mounted the bleed duct is quite short and its penalty effect is negligible. The ram air duct runs from the compressor discharge section to the engine inlet. The volume that this ram air duct displaces is part of the area between the engine and nacelle necessary for passage of the engine's secondary air. Since the turbonator air is returned to the secondary flow, the ram air duct uses only part of the total necessary air volume between the engine and the nacelle required for the secondary air. Therefore there is no penalty assigned to the duct for an engine-mounted turbonator which exhausts into the engine's secondary air.

The turbonator which exhausts directly to ambient is assumed to be located away from the engine. Both the bleed air duct and ram air duct contribute towards a range penalty.

Assuming the same length ducts used in the previous section, the volume of the duct displaces 240 lb of fuel. Since it is quite unlikely that 100% usage of this volume could be realized in fuel, this penalty should be quite pessimistic.

The per-cent decrease in range is obtained by use of the equation just derived which reduces the gross weight by the 960 lb of fuel displaced by the four turbonator systems.

$$\frac{\Delta R}{R} = \left[1 - \frac{\ln \frac{Wf-Wa'}{We}}{\ln \frac{We}{We}}\right] \times 100 = 0.40\%$$

Change in Range Due to Accessory System Cooling

The change in range due to the cooling requirements of the accessory system is based on the penalty imposed on an airplane, which must carry sufficient water to cool the accessories by evaporation. Although water would not be used in actuality, it is a very efficient measure of the cooling load.

When evaluating the change in range, it is assumed that the water is continually used up to the same extent as the fuel until it is completely used at the end of the mission. This water therefore displaces an equal weight of fuel by maintaining the same gross weight and empty weight of the airplane.

The change in weight of the airplane with respect to time may be expressed as follows:

$$\frac{dW}{dt} = -(SFC)F - Q$$

where Q is the rate at which the water is evaporated in lb/hr

$$dW = -(SFC)F dt - Q dt$$

$$\frac{dW}{W} = -\frac{(SFC)F dt}{W} - \frac{Q}{W} dt$$

$$= -\frac{(SFC)}{L/D}\frac{dt}{dR} dR - \frac{Q}{W} dt$$

$$dR = -\frac{V}{SFC}\frac{L}{D}\left(\frac{dW}{W} + \frac{Q}{W} dt\right)$$

$$R = \frac{V}{SFC}\frac{L}{D}\left(\ln \frac{Wf}{We} - \frac{Q dt}{W}\right)$$

By assuming W is the average weight of the airplane, the integral of $Q dt$ is the weight of water.

$$R = \frac{V}{SFC}\frac{L}{D}\left(\ln \frac{Wf}{We} - \frac{Ww}{1/2(Wf+We)}\right)$$

The change in range due to cooling loads may be expressed as

$$\frac{\Delta R}{R} = 1 - \frac{\ln \frac{Wf}{We} - \frac{Ww}{1/2(Wf+We)}}{\ln \frac{Wf}{We}}$$

$$\frac{\Delta R}{R} = -\frac{Ww}{1/2(Wf+We)} \frac{1}{\ln \frac{Wf}{We}}$$

Since the turbonator losses (drive and generator) are carried away by the turbine exhaust air, there is no cooling requirement and therefore no cooling penalty.

The hydraulic constant-speed drive and the generator both have cooling requirements. The losses of these two components is the 67 hp required from the engine multiplied by one minus the product of the drive, generator, and gear efficiencies. These losses result in a total heat dissipation of 55,000 Btu/hr.

One pound of water can absorb approximately 1,000 Btu. Therefore, 55 lb of water are required for each drive and generator per hour of flight.

For the case of the bomber with four generators, 220 lb of water per hour of flight is required. The per-cent decrease in range as derived in the foregoing calculates as

$$\frac{\Delta R}{R} = \left[\frac{Ww}{1/2(Wf+We)} \frac{1}{\ln Wf/We}\right] \times 100$$

$$= 0.12\%/\text{hr}$$

Interceptor-Type Airplane

The evaluation of the penalties imposed on an interceptor-type airplane by accessory systems has more meaning if the comparison is made in terms of thrust or speed rather than range. This type of comparison is more complicated since it is difficult to obtain a common denominator for power

extraction, weight, volume, and cooling requirements.

One method which has some merit is to compare the R/C (rate of climb) ability of the aircraft which takes into account the change in thrust and drag as well as the increase in weight.

The term which shall be compared may be called the R/C factor, and defined as the excess thrust available divided by the gross weight.

$$R/C \text{ factor} = \frac{F-D}{Wf}$$

The value to be used for an aircraft with no accessory system shall be as follows:

F = thrust at maximum reheat, Mach 3, 70,000 feet

D = thrust at minimum reheat, Mach 3, 70,000 feet, which is the value used for the bomber dash mission.

Wf = gross weight with no accessories 40,000 lb

Change in R/C Factor Due to Power Extraction

$$\frac{\Delta R/C}{R/C} = 1 - \frac{(F-D)_a}{(F-D)}$$

where the term $(F-D)_a$ is the value obtained with power extraction.

Turbonator Using Ram Air Exhausting to Ambient

For the case of an interceptor-type aircraft, the change in the R/C factor as explained is equal to one minus the ratio of the excess thrust with ram air drag to the excess thrust with zero ram air. For this calculation the drag term in the numerator is increased by the amount of ram drag.

$$\frac{\Delta R/C}{R/C} = 1 - \frac{(F-D)_a}{(F-D)}$$

The per-cent decrease in this factor by increasing the drag in the numerator term is 1.04%.

Turbonator Using Ram Air Exhausting into Engine Secondary Air

For the case of the interceptor-type airplane, the change in the R/C factor is again determined by the increase in the drag term an amount equal to the decrease in engine thrust due to the increased secondary air flow.

$$\frac{\Delta R/C}{R/C} = \left[1 - \frac{(F-D)_a}{(F-D)}\right] \times 100 = 0.29\%$$

Hydraulic Constant Speed Drive and Generator Using Shaft Power Extraction

The per-cent change in the R/C factor for the interceptor-type airplane is determined by changing the thrust term in the numerator to the value obtained at maximum reheat with 67-hp extraction.

$$\frac{\Delta R/C}{R/C} = \left[1 - \frac{(F-D)_a}{(F-D)}\right] \times 100 = 0.34\%$$

Change in R/C Factor Due to Accessory Equipment Weight

The change in the R/C factor due to equipment weight for an interceptor-type airplane is determined from the change in gross weight of the airplane:

$$\frac{\Delta R/C}{R/C} = 1 - \frac{Wf}{Wfa}$$

The per-cent change in R/C is as follows:

Interceptor with One Turbonator	Change in R/C, Per Cent
Engine-mounted	0.58
Remotely located	0.63
Hydraulic constant-speed drive and generator	0.75

Change in R/C Factor Due to Accessory Equipment Volume

Determining the effect of volume on the aircraft is not as straightforward as the other factors. The major problem concerning this study is the volume of the ducts, since the turbonator and hydraulic constant-speed drive and generator volumes are approximately the same. An interceptor-type plane is usually a high-density airplane. The major penalty of ducts is therefore the finding of available space. Since it must be assumed that no space is available on a high-density airplane, the ducts must displace fuel which actually decreases the gross weight of the airplane since the duct weight is negligible in comparison. However, the weight of fuel must be regained for mission purposes. For this study, it is assumed that the displaced fuel is carried by adding wing tanks.

The drag term in the numerator of the rate of climb factor is increased by the amount of the external fuel tank drag:

$$\frac{\Delta R/C}{R/C} = 1 - \frac{(F-D)_a}{(F-D)}$$

The drag of the external fuel tank is estimated by assuming a drag coefficient of 0.15 based on cross-sectional area and a fineness ratio of 8 to 1. As has been explained, the duct volume penalty applies only for the remotely located turbonator. The duct length for such an installation could run up to 15 feet. The volume of this duct is assumed to displace fuel for the case of a high-density interceptor-type airplane.

The volume penalty as explained is assumed to be the result of the drag of an external tank to hold the displaced fuel.

The 15-foot duct will displace approximately 120 lb of fuel (18.5 gallons). It is estimated that the drag of an external tank for this capacity would be approximately 60 lb, which is the increase in drag of the numerator term:

$$\frac{\Delta R/C}{R/C} = \left[1 - \frac{(F-D)_a}{(F-D)}\right] \times 100 = 1.31\%$$

Change in R/C Factor Due to Accessory System Cooling

The change in the R/C factor due to the cooling requirement is based on the amount of water which must be carried. The penalty is based on the weight of the water as well as the volume type of penalty. The water is assumed to be carried in an external fuel tank resulting in a drag penalty.

$$\frac{\Delta R/C}{R/C} = 1 - \frac{Wf}{Wfa}\,\frac{(F-D)_a}{(F-D)}$$

Since the external tank drag increases nonlinearly with weight of water, the penalty must be evaluated for various times of flight.

The weight and drags assumed for determining the change in the rate of climb factor are listed in the following:

Hours	Ww	D(Tank)	ΔR/C
1	55	30	0.80%
3	165	62	1.78%
5	275	86	2.57%

The water weight increases the gross weight term Wfa and the drag decreases the thrust term $(F-D)_a$:

$$\frac{\Delta R/C}{R/C} = 1 - \frac{Wf}{Wfa}\,\frac{(F-D)_a}{(F-D)}$$

Reference

1. The Turbonator—Design and Development, J. T. Duane, J. W. Harrison. *AIEE Transactions*, vol. 77, pt. II, 1958 (*Paper no. 58-791*).

Effect of Operating Frequency on the Weight and Other Characteristics of Missile Alternators and Transformers

R. E. TURKINGTON
MEMBER AIEE

THE extensive use of 400-cps (cycle-per-second) power systems in manned military aircraft and the consequent availability of standard 400-cps generation, conversion, and utilization components has caused this frequency to be used in most guided missiles currently in production or under development in the United States. However, as higher missile flight speeds, longer ranges, and greater accelerations are needed to achieve successful guided-missile missions, the adoption of all measures that reduce the weight and volume of the electric power system without impairing reliability and logistics must be considered. The use of a power frequency higher than 400 cps would permit substantial reductions in the weight and volume of most missile electromagnetic components and of the prime energy source. However, the determination of an optimum missile power frequency must include a coordinated study[1,2] of the generation, conversion, transmission, and utilization of electric power in missiles.

In this paper, significant characteristics of transformers and alternators (such as weight, current and flux densities, winding resistance, leakage reactance, core loss, and winding loss) are determined as functions of the operating frequency; curves of relative weight versus frequency also are presented. Induction, permanent-magnet, and wound-field alternators, as well as power transformers, are considered.

Paper 58-772, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted June 7, 1957; made available for printing March 21, 1958.

R. E. TURKINGTON, now with the Radio Corporation of America, Airborne Systems Laboratory, Waltham, Mass., was formerly with the Massachusetts Institute of Technology, Cambridge, Mass.

The author gratefully acknowledges the assistance and encouragement given by Prof. D. C. White during the conduct of this study and also his comments and suggestions pertaining to the manuscript.

The work presented in this paper was done at the Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, and was supported by the Bureau of Aeronautics, Department of the Navy, under Contract no. *NOas 54-583-c.*

Method of Analysis

The relationships between weight and frequency for transformers and alternators should be determined with all other independent parameters, such as power rating, output voltage, operating life, ambient temperature, and quality of design, held constant. For this study, the term "quality of design" is defined as the nearness of a design to an optimum, or minimum-weight, design. Obviously, a distorted relationship would result if an optimum 400-cps design were compared with a 4,000-cps design that is far from optimum.

Either of two approaches can be used to determine the weight-frequency characteristics. The study could be quite specific; that is, a series of optimum, or minimum-weight, transformers and alternators could be designed for several power ratings and several operating frequencies. This approach, even with the assistance of a digital computer,[3] would be prohibitively laborious and time-consuming in view of the large number of units that must be designed. Also, because the design of minimum-weight components often is limited by thermal considerations, the design process is largely dependent on empirical methods or data; furthermore, the quality of design might be difficult to evaluate or to control and could be verified only by experimentation.

In this paper, the weight-frequency characteristics for alternators and transformers are determined by considering only gross effects and neglecting many of the details associated with the computation of an actual design. The analyses are based on the postulated existence of an experimentally proved unit (called the model) designed for operation at a particular frequency. This design is scaled analytically while maintaining geometric similarity in order to obtain a generalized unit (called the prototype) designed for operation at an arbitrary frequency. Geometric similarity is maintained by multiplying each linear dimension of the model (except the core-lamination thickness and the conductor diameter, which are treated separately) by a linear scaling factor, or size factor.

The frequency spectrum to be considered is from 400 to 5,000 cps.

Table I. Transformer Characteristics as a Function of Frequency for Constant Induced Volt-Amperes and Constant Temperature

Transformer Characteristics, $M(f) \propto f^{m_f}$	Convective Cooling* Exponent	Convective Cooling* $\frac{M(4,000)}{M(400)}$	Conductive Cooling* Exponent	Conductive Cooling* $\frac{M(4,000)}{M(400)}$
Weight, W_t	$-6\left(\frac{\beta-\alpha}{7\beta-2}\right) = -0.143$	0.720	$-3\left(\frac{\beta-\alpha}{3\beta-2}\right) = -0.217$	0.607
Size factor, s	$-2\left(\frac{\beta-\alpha}{7\beta-2}\right) = -0.048$	0.896	$-\left(\frac{\beta-\alpha}{3\beta-2}\right) = -0.072$	0.847
Flux density, B_{wt}	$-\left(\frac{7\alpha-2}{7\beta-2}\right) = -0.833$	0.147	$-\left(\frac{3\alpha-2}{3\beta-2}\right) = -0.784$	0.164
Current density, J	$\left(\frac{\beta-\alpha}{7\beta-2}\right) = 0.024$	1.06	$\left(\frac{\beta-\alpha}{3\beta-2}\right) = 0.072$	1.18
Number of turns, N	$-3\left(\frac{\beta-\alpha}{7\beta-2}\right) = -0.072$	0.847	$-\left(\frac{\beta-\alpha}{3\beta-2}\right) = -0.072$	0.847
Per-unit resistance, R_{pu}	$-4\left(\frac{\beta-\alpha}{7\beta-2}\right) = -0.096$	0.802	$-\left(\frac{\beta-\alpha}{3\beta-2}\right) = -0.072$	0.847
Per-unit reactance, X_{pu}	$\frac{8\alpha-\beta-2}{7\beta-2} = 0.808$	6.04	$\frac{3\alpha-2}{3\beta-2} = 0.784$	6.08
Conductor area, A_c	$-\left(\frac{\beta-\alpha}{7\beta-2}\right) = -0.024$	0.946	$-\left(\frac{\beta-\alpha}{3\beta-2}\right) = -0.072$	0.847
Winding loss, P_w	$-4\left(\frac{\beta-\alpha}{7\beta-2}\right) = -0.096$	0.802	$-\left(\frac{\beta-\alpha}{3\beta-2}\right) = -0.072$	0.847
Core loss, P_i	$-4\left(\frac{\beta-\alpha}{7\beta-2}\right) = -0.096$	0.802	$-\left(\frac{\beta-\alpha}{3\beta-2}\right) = -0.072$	0.847
Volt-ampere excitation, $(VA)_e$	-0.393	0.405	-0.371	0.425

* Two-mil Hipersil is assumed as core material for the computation of the numerical results.

Table II. Alternator Characteristics as Functions of Frequency for Constant Induced Volt-Amperes

Alternator Characteristics, $M(f) \propto f^{m\alpha}$	Wound-Field Alternator (Case I)* Exponent‡	Wound-Field Alternator (Case I)* $\frac{M(4,000)}{M(400)}$	Wound-Field Alternator (Case II) Permanent-Magnet Alternator† Exponent‡	Wound-Field Alternator (Case II) Permanent-Magnet Alternator† $\frac{M(4,000)}{M(400)}$	Induction Alternator‡ Exponent‡	Induction Alternator‡ $\frac{M(4,000)}{M(400)}$
Weight, W_t	$-3/4$	0.179	$-6\left(\frac{\beta-\alpha}{7\beta-4}\right) = -0.384$	0.414	$-6/7$	0.139
Size factor, a	$-1/4$	0.563	$-2\left(\frac{\beta-\alpha}{7\beta-4}\right) = -0.128$	0.745	$-2/7$	0.518
Flux density, B_{mf}	$-1/8$	0.750	$-\left(\frac{7\alpha-4}{7\beta-4}\right) = -0.552$	0.280	‡	1.00
Stator current density, J_1	1/8	1.33	$\left(\frac{\beta-\alpha}{7\beta-4}\right) = 0.064$	1.16	1/7	1.39
Number of stator turns, N_1	$-1/2$	0.316	$-3\left(\frac{\beta-\alpha}{7\beta-4}\right) = -0.192$	0.643	$-3/7$	0.373
Per-unit resistance, R_{pu}	$-1/2$	0.316	$-4\left(\frac{\beta-\alpha}{7\beta-4}\right) = -0.256$	0.555	$-4/7$	0.268
Per-unit reactance, X_{up}	0	1.000	$\left(\frac{8\alpha-\beta-4}{7\beta-4}\right) = 0.488$	3.08	$-1/7$	0.720
Stator conductor area, A_{c1}	0	1.000	$-\left(\frac{\beta-\alpha}{7\beta-4}\right) = -0.064$	0.863	$-1/7$	0.720
Stator winding loss, P_{w1}	$-1/2$	0.316	$-4\left(\frac{\beta-\alpha}{7\beta-4}\right) = -0.256$	0.555	$-4/7$	0.268
Stator core loss, P_{i1}	0.245	1.76	$-2\left(\frac{\beta-\alpha}{7\beta-4}\right) = -0.127$	0.746	0.393	2.47
Stator core temperature, Δt_{m1}	0.525	3.35	†	1.00	0.679	4.78
Magnetizing current, I_ϕ					1/7	1.39
Excitation capacitance, C					$-6/7$	0.139
Capacitor volume, V_c					$-6/7$	0.139

* Stator-winding and field-winding temperatures maintained constant.

† Stator-winding and stator-core temperatures maintained constant.

‡ Stator-winding temperature and flux density maintained constant

‡Five-mil Tran-Cor T is assumed for the evaluation of those exponents written in decimal form. Exponents written in fractional form are not a function of the core material.

Table III. Core-Material Constants for Use With Equations 2, 22, and 24

Material	Thickness, Mils	Core-Material Constants $\gamma^* \times 10^{13}$	α	β	$\nu^* \times 10^{15}$	σ	λ
Hipersil	2	0.15	1.68	1.96	12.5	1.34	1.91
	5	52.2	1.36	1.69	76.5	1.37	1.67
	12		{1.38}	{1.69}			
			{1.44}	{2.48}			
Monimax	4		{1.43}	{1.88}			
			{1.61}	{1.95}			
Sinimax	4		{1.35}	{1.76}			
			{1.57}	{1.83}			
Tran-Cor T	5	60.2	1.25	1.80			
Silectron T	7	23.3	1.36	1.72			
USS Transformer 72	14	6.2	1.52	1.91			

* These values of γ and ν are valid when p_t is in watts per cubic inch, $(v_a)_s$ is in volt-amperes per cubic inch, and B_{mf} is in gauss.

Results of Analysis

The functional relationships between frequency and the variables that characterize the design of transformers and alternators are derived in Appendixes I through IV and are summarized in Tables I and II. These tables give values for the exponent in the expression

$$M(f) \propto f^{m\alpha p} \qquad (1)$$

where $M(f)$ is a general variable or characteristic and f is the operating frequency; values of the ratio $M(4,000)/M(400)$ are also given. The constants α and β are characteristics of the core material and are defined by the relationship

$$p_t \propto f^{\alpha} B_{mf}^{\beta} \qquad (2)$$

where

p_t = core loss per unit volume

B_{mf} = the maximum value of the core flux density

Values of α and β for several core materials are given in Table III.

In order to obtain relatively simple, general relationships of the type shown as equation 1, the induced voltage and the induced volt-amperes, rather than the output voltage and the output volt-amperes, have been maintained constant.

Curves of relative weight of the electromagnetic portions of transformers and alternators as functions of frequency for constant volt-ampere output and constant output voltage are shown in Figs. 1 and 2. Unity relative weight is shown for 400-cps units. The weights shown in the figures are greater than those indicated in Tables I and II because of the voltage drop across the output-winding resistance and leakage reactance, both of which are frequency dependent.

Because the assumptions on which these analyses are based have a marked effect on the weight-frequency relationships, a more detailed discussion of the results, including a statement of the major thermal assumptions, follows. The thermodynamic characteristics assumed for the model transformer and the model alternators are discussed in the appendixes

Transformers

The relative weight of power transformers as a function of frequency is shown in Fig. 1. Both curves are computed for constant volt-ampere output and constant core and winding temperatures. For curve *1*, cooling of the transformer by either forced or free convection is assumed; for ease of analysis, the surface temperature of the winding and core are maintained constant. For curve *2*, cooling of the transformer by thermal conduction is assumed; both the maximum internal winding and core temperatures are maintained constant. These curves have been plotted from equations 37 and 40 in Appendix I.

Wound-Field Synchronous Alternators

Curve *1* in Fig. 2 shows the relative weight of wound-field synchronous alternators as a function of frequency when the volt-ampere rating, the maximum internal temperature, and the number of poles are maintained constant.

The thermal criteria which limit weight reduction of 400-cps aircraft alternators often are the field-winding and the armature-winding temperatures; the stator core temperature often is below the maximum allowable value. Therefore, for the lower frequency segment of curve *1*, the temperature of both the field winding and stator winding are assumed constant and equal to the maximum allowable value; the temperature rise of the stator iron is assumed equal to one half the maximum allowable rise. (This is designated as

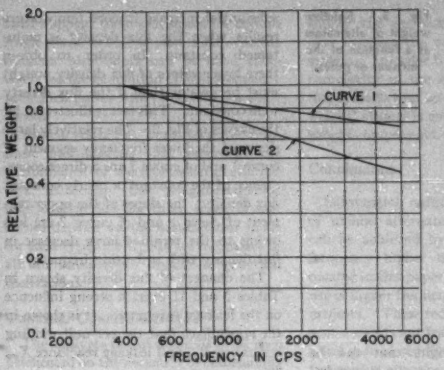


Fig. 1. Relative weight of transformers as a function of frequency

Curve 1—Convectively cooled transformer
Curve 2—Conductively cooled transformer

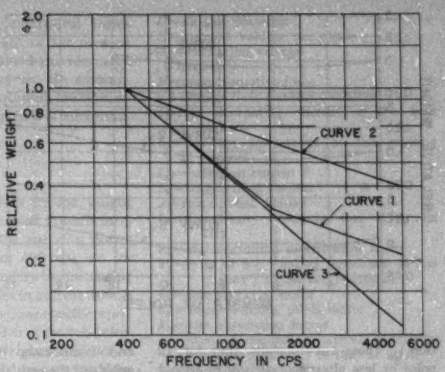


Fig. 2. Relative weight of alternators as a function of frequency

Curve 1—Wound-field synchronous alternator
Curve 2—Permanent-magnet alternator
Curve 3—Induction alternator

case I in Table II.) As the frequency is increased from 400 cps, the temperature of the stator iron increases owing to the increase in stator core loss and reaches the maximum permissible value at 1,500 cps. For further increases in frequency, the temperature of both the stator winding and the stator iron are assumed constant (case II in Table II). During this interval, the field temperature decreases. The lower frequency and higher frequency segments of curve *1* have been plotted from equations 84 and 86, respectively, in Appendix III. The slope of the curve below 1,500 cps is not dependent on the core-loss characteristics of the core material.

Permanent-Magnet Alternators

The relative weight of permanent-magnet synchronous alternators as a function of frequency when the volt-ampere rating, the stator winding and core temperatures, and the number of poles are maintained constant is shown by curve 2 of Fig. 2. Because the stator winding and core temperatures have been assumed constant, curve *2* has essentially the same slope as the higher frequency portion of curve *1*, and the relationships between the various alternator variables and frequency for constant induced volt-amperes are the same as for the wound-field alternator with constant stator winding and core temperatures (see Table II).

Induction Alternators

Curve *3* of Fig. 2 shows the relative weight of induction alternators as a function of frequency for constant volt-ampere output, stator winding temperature, flux density, and constant number of poles. The core flux density is assumed constant for two reasons: 1. Capacitively excited, isolated induction alternators normally operate with their cores in a saturated state[4] in order to provide a stable output.* 2. Because the total excitation-capacitor volume is a function of the product of the core flux density and the saturation factor, as shown by equation 66 in Appendix II, and because the saturation factor increases rapidly as the flux density is increased above the saturation value, the core should not be highly saturated in a 400-cps unit if excessive physical size of the capacitors is to be avoided.

Core loss, therefore, cannot be controlled by the usual technique of decreasing the core flux density, and, as indicated in Table II, significant core-temperature variations may result for a given type and thickness of core material. However, the core temperature may be kept nearly constant over a considerable frequency range by using more effective cooling methods and by progressing from relatively thick core laminations to lower core-loss material as higher frequencies are considered. Because only limited improvements in the temperature-frequency relationship can thus be achieved, a maximum frequency may exist above which induction alternators may not be practical. However, this maximum frequency is dependent on many factors, including power rating, total operating life, and duty cycle and cannot be determined on a generalized basis; for a continuous-duty 1.2-kw 60,000-rpm induction alternator,[4] the maximum frequency is well above 2,500 cps.

Effect of Varying the Number of Poles

In order to maintain a constant ratio between the tooth and the yoke flux density and, therefore, to make possible the derivation of generalized curves of weight versus frequency, the number of poles has been assumed constant. However, because factors such as prime-mover efficiency, bearing characteristics, and strength of materials prevent the design of both 400-cps and 5,000-cps alternators with the same number of poles, the relationship between weight and the number of poles should be considered. Three curves of relative total weight as a function of the number of poles for constant output power and constant shaft speed are shown in Fig. 3. Curve *1* (from unpublished notes by W. H. Gable) is computed for a 4.5-kw 24,000-rpm induction alternator; curve *2* is for a 10-kva 12,000-rpm synchronous alternator; and curve *3* is for a 6,000-rpm synchronous alternator. (Curves *2* and *3* are based on data from Westinghouse Electric Corpora-

* Methods which permit stable operation with nonsaturated cores include the following. 1. Operation with saturated reactors placed in series with the excitation capacitors.[4] In this way, high core loss and the associated high temperatures occur in the reactors. 2. Use of a regulator designed to provide stable operation of an inherently unstable alternator.

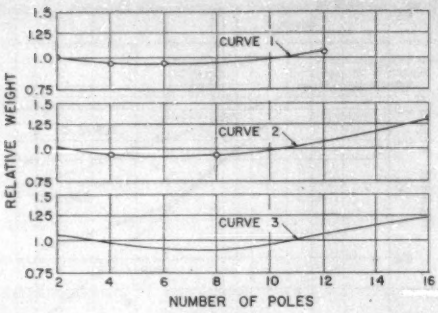


Fig. 3. Relative weight of alternators as a function of the number of poles

tion.) These curves show that for 12 poles or less alternator weight is almost independent of the number of poles. This result correlates well with the assumed mathematical model: Equation 55 in Appendix II shows that the induced volt-amperes are proportional to the ratio of frequency to the number of poles (that is, proportional to the synchronous speed) and are not a function of either frequency or number of poles separately. The differences between Fig. 3 and the mathematical model are due mainly to the decreasing volume of end turns as the number of poles is increased from two, the possible necessity of increasing the air-gap diameter with increasing number of poles in order to provide sufficient space for the additional poles and slots, and the impossibility of maintaining both geometric similitude and the same degree of saturation in both the stator teeth and stator yoke when the number of poles is changed.

Selection of Core Material

The selection of a core material is essential if numerical results or a curve of weight versus frequency is desired. This study has made no attempt to determine the best core material for high-frequency equipment. Hipersil,[8] a grain-oriented silicon steel, and Tran-Cor T,[8] a nonoriented material, have been selected for use in this study because of their availability in relatively thin gages and because their core-loss and excitation characteristics can be approximated closely by equations 22 and 24 in Appendix I.

Discussion of Results

High-Frequency Operation

The generalized curves of weight versus frequency shown in Figs. 1 and 2 are indicative of the reductions in alternator and transformer weight that can be achieved through the use of a power frequency higher than 400 cps. The actual weight reductions that can be realized in practice depend on factors such as the particular power level, method of cooling, type of alternator, and type of prime mover, and, therefore, must be determined for each particular application.

Because significant reductions in alternator weight can be achieved only by increasing the alternator speed, as indicated by Fig. 3 and equation 55 in Appendix II, the maximum practical shaft speed, rather than the frequency, may limit the magnitude of the weight reduction. Consider, for example, the weight of a 3,200-cps 500-watt induction alternator driven by a hot-gas turbine relative to the weight of a corresponding 400-cps 2-pole alternator. Because the maximum practical shaft speed for a short-life unit of this size and type is approximately 100,000 rpm, a shaft speed of 96,000 rpm is assumed. This speed corresponds to a 2-pole 1,600-cps alternator, the weight of which, relative to that of the 400-cps unit, is 0.29 from curve *3* of Fig. 2. The desired frequency, 3,200 cps, is obtained by doubling the number of poles; this permits a further weight reduction of 0.95, as shown by curve *1* of Fig. 3. The relative weight of the 3,200-cps alternator, therefore, is 0.29×0.95 or 0.28. For comparison consider the generation of 9,600 cps by increasing the number of poles to 12; from Figs. 2 and 3, the weight of this alternator relative to the 400-cps unit is 0.29×1.05 or 0.30.

Tables I and II illustrate some effects of high-frequency operation of electromagnetic energy-conversion devices. Because the core loss per unit volume of the core material increases rapidly with increasing frequency, a substantial decrease in flux density is required to maintain constant core temperature; conversely, a substantial increase in core temperature results when the flux density is maintained constant. In order to obtain these lower values of flux density, weight must be greater than if the flux density were constant. This fact is illustrated by the curves of Fig. 2. The relatively large slopes of the lower frequency segment of curve *1* and of curve *3* are a direct consequence of the constant or nearly constant flux density; the slopes of the upper segment of curve *1* and of curve *2* are less owing to the required large decrease in flux density with increasing frequency.

The changes in flux density shown in Tables I and II exert a strong influence on the leakage reactance. It is shown in the appendixes that the per-unit winding resistance R_{pu} and leakage reactance X_{pu} are

$$R_{pu} \propto \frac{J}{B_{m1}} \frac{1}{af} \tag{3}$$

and

$$X_{pu} \propto \frac{J}{B_{m1}} a \tag{4}$$

where

J = the rms current density
a = the size factor

Therefore, when B_{m1} must be decreased in order to maintain constant temperature, X_{pu} increases rapidly; when B_{m1} is constant or nearly constant, X_{pu} is either constant or may even decrease with increasing frequency. The maximum voltage drop across the output-winding resistance and leakage reactance for the conditions covered in Figs. 1 and 2 is approximately 14% of the induced voltage.

Other observations that can be made from Tables I and II are

1. The winding, or copper, loss for each device decreases with increasing frequency.
2. The current density must be increased with frequency in each case to maintain constant winding temperature.
3. The total core loss of alternators and transformers decreases with increasing frequency when the core temperature is maintained constant.
4. The volt-ampere excitation required by the transformer core decreases with increasing frequency.
5. The induction-alternator magnetization current increases but the capacitance and volume of the excitation capacitors decrease substantially with increasing frequency.

Other Types of Alternators

Additional alternator types that merit consideration are the flux switch alternator[9] and the variable-reluctance, or inductor, alternator. Flux-switch alter-

nators appear to be well suited for certain single-phase missile applications, whereas inductor alternators can be used effectively for the generation of high-frequency prelaunch power with relatively low shaft speeds.

The curves of Fig. 2 are computed from basic relationships that described the performance of electromagnetic alternators. Although the weight-frequency relationships for the three types of machines depend on the assumptions made for each type, the trends are indicative of all alternator types capable of practical realization.

Thermal Considerations

The analyses in this paper have all been predicated on the necessity of maintaining constant steady-state winding or core temperatures or both. Consequently, thermal considerations have determined the maximum winding and/or core loss. However, in applications where missile generators or transformers have a very short tactical operating life, perhaps only 15 seconds, these devices never reach a thermal steady state. Equipment designed to take advantage of this transient, of intermittent-duty, operation usually is substantially lighter than equivalent steady-state components. Consider the core loss and core heating; if all the heat generated in the core is assumed to be stored in the core and the heat transferred away from the core is neglected, the maximum core loss per unit volume can be approximated as[4]

$$p_t = 3.75 \frac{\Delta t_{mt}}{T} \qquad (5)$$

where

p_t = the specific core loss in watts per pound (w/lb)
Δt_{mt} = the maximum core temperature rise in degrees centigrade (C)
T = the operating time in minutes (min)

Thus, if a temperature rise of 240 C is desired after an operating period of 5 min, the specific core loss must be 180 w/lb, a value that corresponds approximately to 5-mil Tran-Cor T operated at 10,000 gauss and 5,000 cps. The maximum permissible winding loss and current density can be determined similarly.

Because minimum-weight electric systems may not always be achieved with minimum-weight components, intermittent-duty devices are difficult to treat on a generalized basis. The magnitude of the losses indicated by equation 5 for a particular component may be so great that the increase in the weight of the prime-energy source and the associated energy-conversion equipment required to supply the increased component losses may increase the weight of the over-all system. In order to avoid this increased complexity in the analysis, the assumption is made that an optimum (minimum-weight) system is comprised of minimum-weight components.

Conclusions

Substantial reductions in the weight of missile alternators and transformers can be achieved by the use of a power frequency higher than 400 cps, as indicated in this paper and by the increasing use of higher frequencies in current missile projects. These reductions usually represent a significant fraction of the total electrical power-supply weight: 31% for a typical 400-cps solid-fuel-powered turbine-alternator supply designed for an air-to-air missile if 3,200 cps were adopted. A study[1,2] of missile utilization components indicates that higher power frequencies are compatible with missile utilization components and that the weight of many electromagnetic utilization components may be decreased substantially.

The use of a higher power frequency also can result in decreased weight of the prime power source, defined here as the prime energy source (solid or liquid fuel or pressurized gas) and a partial-admission turbine. Turbine efficiency is approximately proportional to turbine tip speed over a wide range of speed.[3] Therefore, the higher alternator speeds that can be used with higher frequencies permit a decrease in the weight of the fuel-turbine system. The fourfold increase in speed assumed for the example of the preceding section allows an estimated weight reduction of the prime power source of 19% of the total 400-cps power-supply weight. The weight reduction for the entire hypothetical 3,200-cps supply is estimated at 65%. (The hypothetical supply consists of a solid-fuel-powered turbine, an induction alternator, a transformer-rectifier-filter unit, and a suitable regulating device.)

The weight reductions that can be achieved with high-frequency high-speed power supplies are so significant that missile engineers interested in reducing missile weight must take cognizance of this fact in future designs.

Nomenclature

a = size factor, or scaling factor
A_i = gross core cross-sectional area
A_s = exposed surface area (effective heat-transfer area)
A_w = gross winding cross-sectional area
B_m = maximum flux density
C = excitation capacitance
D = air-gap diameter
E = induced voltage per phase
E_f = excitation voltage
f = frequency
F = magnetomotive force
h = average heat-transfer coefficient
I = rms alternating current or d-c field current
I_c = core-loss component of the exciting current
I_e = excitation current
I_ϕ = magnetizing current
J = rms current density
k_b = flux-density constant (see Appendix III)
k_d = stator-winding distribution factor
k_f = ratio of a-c winding resistance at frequency f to the d-c resistance
k_i = core stacking factor
k_p = stator-winding pitch factor
k_s = core saturation factor
k_t = ratio of conductor-material resistivity at temperature t to that at 20 C
k_w = winding space factor
k_τ = ratio of average stator tooth width to the air-gap tooth pitch
K = mmf proportionality constant, defined by equations 73, 75, and 77
K_e = voltage constant; $K_e = 8.88 k_p k_d k_b k_i \tau$ ($k_b = 0$ for the induction alternator)
K_p = permeance factor, defined by equations 15 and 52
K_t = thermal conduction constant, defined by equations 30, 31, and 61
K_L = field-winding loss constant, defined by equation 80
K_ϕ = magnetizing current constant; $K_\phi = \pi k_f k_\tau / 2\sqrt{2}\mu_0 k_p k_d$
l = axial length of the core
l_i = average length of the flux path within the core
l_w = mean length of winding turn
L = leakage inductance per phase
N = number of series turns per phase
p = number of poles
p_t = core loss per unit volume
P = total loss
q = number of phases
R = winding resistance per phase
R_c = resistance representing core loss
s = slip
t_m = maximum temperature
t_s = average surface temperature
T_m = maximum allowable temperature
V = terminal voltage
$(\text{vo})_e$ = excitation volt-amperes per unit volume
$(VA)_e$ = total excitation volt-amperes
$(VA)_i$ = total induced volt-amperes
$(VA)_t$ = total output volt-amperes
$(VA)_\phi$ = total magnetizing volt-amperes
W_t = weight in terms of $(VA)_t$
W_i = weight in terms of $(VA)_i$
X = leakage reactance per phase
X_ϕ = magnetizing reactance
Z_{bk} = base impedance referred to the kth winding
Z_L = load impedance
$\alpha, \beta, \gamma, \lambda, \nu, \sigma$ = core-material constants defined by equations 2, 22, and 24
θ = load power-factor angle
μ_0 = permeability of free space
ρ = conductor-material resistivity
ρ_e = effective resistivity; $\rho_e = k_f k_t \rho$
Φ_m = maximum flux

Subscripts

o = refers to a characteristic of the model
1 = refers to transformer primary or alternator stator

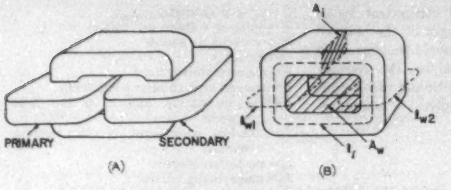


Fig. 4. Single-phase core-type transformer

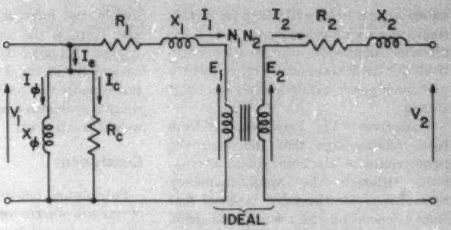


Fig. 5. Transformer equivalent circuit

s = refers to transformer secondary or alternator rotor
a = denotes units of length
g = refers to the air gap
i = refers to the core (except as previously noted)
pu = per-unit quantity
w = refers to the winding

Appendix I. Power Transformers

The derivations in this appendix can be applied equally well to the single-phase core-type transformer shown in Fig. 4(A), to the shell type, or to the "simple" type; if suitable changes in proportionality constants and correct definitions of the physical parameters are made, the relationships also can be applied to 3-phase transformers. The primary winding may be placed on one leg and the secondary on the other, as shown in Fig. 4(A), or one half of the primary and one half of the secondary may be placed on each leg. In Fig. 4(B), certain physical parameters are defined: A_f is the cross-sectional area of the core and is assumed constant in all parts of the core, l_f is the mean length of the flux path in the core, A_w is the window area, and l_{w1} and l_{w2} are the mean lengths of turn for the primary and secondary windings, respectively. The assumptions inherent in the formulation of the often-used equivalent circuit of Fig. 5 are made for this analysis.

Basic Relationships

Induced Voltage and Current

If the flux varies sinusoidally with time, the steady-state primary and secondary induced voltage are

$$E_k = 4.44 N_k f \Phi_{mf}, \qquad k = 1,2 \tag{6}$$

where

E_1, E_2 = the rms primary and secondary induced voltages
N_1, N_2 = the number of primary and secondary turns
f = the frequency
Φ_{mf} = the maximum value of the core flux

Any consistent system of units can be used with this and the following equations. If the meter-kilogram-second system is used (Φ expressed in webers), E is given in volts. The maximum flux is

$$\Phi_{mf} = k_f A_f B_{mf} = k_f A_w B_{mf} a^2 \tag{7}$$

where

k_f = the core space factor, or stacking factor
A_f = the gross core cross-sectional area of the prototype
B_{mf} = the maximum core flux density
a = the size factor, or scaling factor

and the subscript o is used to designate the geometry of the model. The constant k_f is defined as the ratio of the actual, or net, iron cross-sectional area to the gross core cross-sectional area. If equation 7 is substituted into equation 6, the induced voltages become

$$E_k = 4.44 k_f A_{fo} f N_k B_{mf} a^2, \qquad k = 1,2 \tag{8}$$

If the window area is divided equally between the primary and secondary windings, the conductor cross-sectional area can be written as

$$A_{ck} = \frac{k_w A_w}{2N_k} = \frac{k_w A_{wo} a^2}{2N_k}, \qquad k = 1,2 \tag{9}$$

where

A_{c1}, A_{c2} = the cross-sectional area of the primary and secondary conductors
k_w = the winding space factor
A_w = the window area

The constant k_w is defined as the ratio of the total area of conductor material passing through the window to the total window area. For simplicity, the space factor is assumed equal for both the primary and secondary windings.

If the excitation current is assumed small compared with the primary current and if the primary and the secondary rms current densities both are assumed equal to J, the rms primary and secondary currents I_1 and I_2 are

$$I_k = J A_{ck} = \frac{k_w A_{wo}}{2} \frac{J a^2}{N_k}, \qquad k = 1,2 \tag{10}$$

Winding Impedances

It is convenient to define a base impedance as

$$Z_{bk} \equiv \frac{E_k}{I_k} = \frac{8.88 k_f A_{fo}}{k_w A_{wo}} \frac{f N_k^2 B_{mf}}{J}, \qquad k = 1,2 \tag{11}$$

where Z_{b1} and Z_{b2} are the base impedance when referred to the primary and the secondary, respectively.

The primary and secondary resistances R_1 and R_2 can be written as

$$R_k = \frac{2 \rho_s l_{wko}}{k_w A_{wo}} \frac{N_k^2}{a}, \qquad k = 1,2 \tag{12}$$

where

l_{w1}, l_{w2} = the mean length of turn of the primary and secondary
$\rho_s = k_t k_f \rho$, the effective resistivity at the average winding temperature t
k_t = the ratio of the resistivity at temperature t to that at 20 C
k_f = the ratio of the a-c resistance of the winding at frequency f to the d-c resistance
ρ = the d-c resistivity

If l_{w1} and l_{w2} are assumed equal, R_1 and R_2 are equal when referred to the same winding, as are the per-unit resistances R_{pu1} and R_{pu2}. From equations 11 and 12,

$$R_{puk} \equiv \frac{R_k}{Z_{bk}} = \frac{\rho_s l_w}{4.44 k_t A_w} \frac{J}{f B_{mf} a}, \qquad k = 1,2 \tag{13}$$

where $l_w = l_{w1} = l_{w2}$.

Equations for the total leakage inductances[8,10] of various winding arrangements have the form

$$L_{lk} = 4\pi \mu_0 N_k^2 G, \qquad k = 1,2 \tag{14}$$

where

L_{l1}, L_{l2} = the total leakage inductance referred to the primary or the secondary, respectively
μ_0 = the permeability of free space
G = a function of geometry

The function G generally is comprised of several terms of the form A/l where A is the cross-sectional area of a leakage-flux path and l is the length of the path external to the iron core. Because geometric similitude has been assumed, $G = aG_o$. Therefore, a permeance factor K_{pa} is defined such that

$$L_k = K_{pa} N_k^2 \qquad k = 1,2 \tag{15}$$

where L_1 and L_2, the leakage inductances of the primary and secondary windings, are assumed equal when referred to the same winding. (The subscript a is used to indicate that K_{pa} has the units of length.) From equation 15, the corresponding leakage reactances are

$$X_k = 2\pi f L_k = 2\pi K_{pao} f N_k^2 a, \qquad k = 1,2 \tag{16}$$

The per-unit reactances are

$$X_{puk} \equiv \frac{X_k}{Z_{bk}} = \frac{\pi k_w A_{wo} K_{pao}}{4.44 k_f A_{fo}} \frac{Ja}{B_{mf}}, \qquad k = 1,2 \tag{17}$$

Volt-Ampere Output

The induced volt-amperes $(VA)_t$ are given by the product of equations 8 and 10:

$$(VA)_t = E_k I_k = 2.22 k_f k_w A_{tc} A_{wc} f J B_m a^4, \quad k = 1,2 \quad (18)$$

If $E_2 \approx V_2$, the secondary induced voltage can be written as

$$E_2 \approx V_2 \left| \frac{1}{\theta} + R_{ps2} + jX_{ps2} \right| \quad (19)$$

where θ is the load power-factor angle. The volt-ampere output $(VA)_t$, therefore, is

$$(VA)_t = V_2 I_2 \approx \frac{(VA)_t}{\left| \frac{1}{\theta} + R_{ps2} + jX_{ps2} \right|} \quad (20)$$

Losses and Excitation

The total winding loss P_w of the transformer is given by

$$P_w = I_k^2 \left[\left(\frac{N_k}{N_1} \right)^2 R_1 + \left(\frac{N_k}{N_2} \right)^2 R_2 \right] = \rho_k k_w l_{wc} A_{wc} J^2 a^3, \quad k = 1,2 \quad (21)$$

In the absence of saturation, the core loss per unit volume p_t is assumed approximated adequately by the expression

$$p_t = \gamma f^\alpha B_m^\beta \quad (22)$$

where γ, α, and β are constants depending on the particular type of core material. Table III shows values of these constants for three grain-oriented materials (Hipersil, Monimax, and Sinimax) and three non-oriented materials (Tran-Cor T, Silectron T, and USS Transformer 72). Fig. 6(A), which shows curves of core loss per unit volume computed from equation 22 and from published Hipersil characteristics,[5] illustrates the range of validity of equation 22. Because many 400-cps designs tend to be temperature-limited rather than excitation-limited, the deviations above the saturation level are not considered detrimental.

The total core loss P_t of the prototype is

$$P_t = k_f l_{tc} A_{tc} \gamma f^\alpha B_m^\beta a^3 \quad (23)$$

The following approximation for the volt-ampere excitation per unit volume of the core $(va)_e$ is assumed:

$$(va)_e = \nu f^\sigma B_m^\lambda \quad (24)$$

Values of ν, σ, and λ for 2-mil and 5-mil Hipersil are given in Table III. Fig. 6(B) compares the approximation represented by equation 24 with data taken from the Hipersil catalog. The total volt-ampere excitation $(VA)_e$ is

$$(VA)_e = k_f l_{tc} A_{tc} \nu f^\sigma B_m^\lambda a^3 \quad (25)$$

Heat Transfer

A detailed treatment of heat transfer in electromagnetic devices is beyond the scope of this study. However, the assumption of geometric similitude between the model and the prototype permits the derivation of relationships between power loss and temperature that involve certain proportionality constants that need not be defined precisely in terms of the transformer dimensions and material properties, provided their functional dependence on the size factor can be determined. Both air cooling (either free or forced convection) and conductive cooling[11] are considered. Solution for the weight-frequency characteristic of an air-cooled transformer is simplified by using the surface temperature as the thermal criterion; for conductive cooling, either the hot-spot or the average internal temperature is used. Heat transfer between windings and core is assumed negligible.

The winding and core losses of an air-cooled transformer must be dissipated from the exposed surfaces of the winding and the core, respectively; thus

$$P_w = h_w A_{ws} \Delta t_{ws} = h_w A_{wso} a^2 \Delta t_{ws} \quad (26)$$

and

$$P_t = h_t A_{ts} \Delta t_{ts} = h_t A_{tso} a^2 \Delta t_{ts} \quad (27)$$

where

h_w, h_t = the average heat-transfer coefficient for the winding and the core, respectively

A_{ws}, A_{ts} = the exposed surface area of the winding and the core

Δt_{ws}, Δt_{ts} = the average surface temperature rise of the winding and of the core above the cooling-air or the ambient temperature

If equations 21 and 23 for P_w and P_t are substituted into equations 26 and 27, expressions for the temperatures are obtained:

$$\Delta t_{ws} = \frac{\rho_k k_w l_{wc} A_{wc}}{h_w A_{wso}} J^2 a \quad (28)$$

and

$$\Delta t_{ts} = \frac{\gamma k_f l_{tc} A_{tc}}{h_t A_{tso}} f^\alpha B_m^\beta a \quad (29)$$

For a conductively cooled transformer, the core loss is assumed to be conducted to the transformer base plate which is mounted to a heat sink. The heat generated in the windings is assumed to be conducted to the base plate and to the heat sink by means of copper or silver thermal conductors inserted in the windings.[11] The expression relating loss and temperature rise is comprised of several terms of the form kA/l, where k is a thermal conductivity, A is the cross-sectional area of a heat-flow path, and l is the length of the path. These terms, or groups of terms are represented by K_{tsw} and K_{tst} for the winding and the core, respectively. Thus, the loss-temperature relationships can be written

$$P_w = K_{tsw} \Delta t_{mw} = K_{tswo} a \Delta t_{mw} \quad (30)$$

and

$$P_t = K_{tst} \Delta t_{mt} = K_{tsto} a \Delta t_{mt} \quad (31)$$

whence

$$\Delta t_{mw} = \frac{\rho_k k_w l_{wc} A_{wc}}{K_{tswo}} J^2 a^2 \quad (32)$$

and

$$\Delta t_{mt} = \frac{\gamma k_f l_{tc} A_{tc}}{K_{tsto}} f^\alpha B_m^\beta a^2 \quad (33)$$

where t_{mw} and t_{mt} are the maximum (or the average, depending on the definition of K_{tsw} and K_{tst}) temperature of the winding and core, respectively.

Weight

Because geometric similarity is maintained, the average density of the core and the windings, including thermal conductors and potting or impregnating material, is assumed independent of the size factor. Furthermore, the weight of the mounting brackets and the case, if one is used, is assumed to be proportional to the total weight of the windings and the core. Therefore, the total transformer weight W can be expressed as

$$W \propto a^4 \quad (34)$$

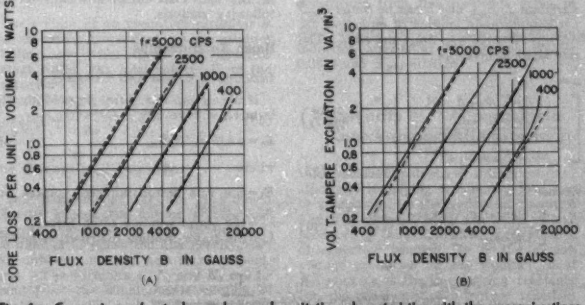


Fig. 6. Comparison of actual core-loss and excitation characteristics with the approximations of equations 22 and 24 for 5-mil Hipersil

A—Core loss per unit volume
B—Excitation per unit volume
Solid curves represent catalog data; dashed curves represent approximations

The Generalized Prototype

The generalized prototype is obtained by scaling the model transformer with the output voltage, the volt-ampere output, and the winding and core temperatures maintained constant; the ambient conditions, including the cooling-air and the heat-sink temperatures, also are assumed constant.

Solution for Convective Cooling

The design of the prototype transformer is specified completely by the five variables N_1, N_2, J, B_{mt}, and a. The variables J, B_{mt} and a are found from the simultaneous solution of equations 18, 28, and 29 for $(VA)_t$, Δt_{tw}, and Δt_{it}; N_2 can be found from equations 8 and 19 for E_2; N_1 can be found from E_2 and V_1.

The size factor is given as

$$a = \left[\frac{(VA)_t^{\beta}(\gamma k_d f_{tw} A_{tu})\times (\rho_{cw} k_w A_{wo})^{\beta/2}}{(2.22 k_{dw} k_w A_{tu} A_{wo})^{\beta}(k_t \times A_{itw}\Delta t_{it})(k_w A_{wo}\Delta t_{tw})^{\beta/2}}\right]^{\frac{2}{7\beta-2}} f^{-2\left(\frac{\beta-\alpha}{7\beta-2}\right)} \quad (35)$$

from equation 34, the weight can be written

$$W_t \propto \left[\frac{\gamma(VA)_t^{\beta}}{\Delta t_{it}(\Delta t_{tw})^{\beta/2}}\right]^{\frac{4}{7\beta-2}} f^{-6\left(\frac{\beta-\alpha}{7\beta-2}\right)} \quad (36)$$

where the subscript t indicates that W_t is a function of the induced volt-amperes $(VA)_t$. The functional relationships between various transformer variables and frequency for constant $(VA)_t$, i_{st}, and i_{tw} are summarized in Table I.

If $(VA)_t$ from equation 20 is substituted into equation 36, weight W_t as a function of the volt-ampere output is obtained:

$$W_t \propto \left[\frac{\gamma(VA)_t^{\beta}}{\Delta t_{it}(\Delta t_{tw})^{\beta/2}}\left|\frac{1}{\theta}+R_{pu2}+jX_{pu2}\right|^{\beta}\right]^{\frac{4}{7\beta-2}} f^{-6\left(\frac{\beta-\alpha}{7\beta-2}\right)} \quad (37)$$

Solution for Conductive Cooling

If conductive cooling is assumed, the variables J, B_{mt}, and a can be found from the simultaneous solution of equations 18, 32, and 33 for $(VA)_t$, Δt_{tw}, and Δt_{it}. The size factor is given as

$$a = \left[\frac{(VA)_t^{\beta}(\gamma k_d f_{tw} A_{tu})\times (\rho_{cw} k_w A_{wo})^{\beta/2}}{(2.22 k_{dw} k_w A_{tu} A_{wo})^{\beta}(K_{ttw}\times \Delta t_{it})(K_{twwo}\Delta t_{tw})^{\beta/2}}\right]^{\frac{1}{3\beta-2}} f^{-\left(\frac{\beta-\alpha}{3\beta-2}\right)} \quad (38)$$

whence

$$W_t \propto \left[\frac{\gamma(VA)_t^{\beta}}{\Delta t_{it}(\Delta t_{tw})^{\beta/2}}\right]^{\frac{2}{3\beta-2}} f^{-3\left(\frac{\beta-\alpha}{3\beta-2}\right)} \quad (39)$$

and

$$W_t \propto \left[\frac{\gamma(VA)_t^{\beta}}{\Delta t_{it}(\Delta t_{tw})^{\beta/2}}\times \left|1/\underline{\theta}+R_{pu2}+jX_{pu2}\right|^{\beta}\right]^{\frac{2}{3\beta-2}} f^{-3\left(\frac{\beta-\alpha}{3\beta-2}\right)} \quad (40)$$

Table I summarizes the functional relationships between various transformer variables and frequency.

Results

Equations 37 and 40, which were used to plot the weight-frequency curves of Fig. 1, contain R_{pu2} and X_{pu2}, both of which are functions of frequency. These terms can be written as

$$R_{pu2} \propto k_f(f)f^{\exp} \quad (41)$$

and

$$X_{pu2} \propto f^{\exp} \quad (42)$$

where the exponents are given in Table I. In order to evaluate the term $(1/\underline{\theta}+R_{pu2}+jX_{pu2})$, typical values of k_f and the proportionality constants for these equations have been determined from measurements on a 3-phase 400-cps 500-watt transformer; $R_{pu2}=0.0158$ and $X_{pu2}=0.00269$ at 400 cps.

Appendix II. Induction Alternators

Conventional induction-motor equivalent circuits can be used to represent a balanced polyphase induction alternator supplying a balanced load, as shown in Fig. 7. Two elements of this circuit are peculiar to alternator applications. First, because the shaft speed is greater than the synchronous speed, the slip s is negative; the negative resistor $[(1-s)/s]R_2$ represents the input shaft power minus the friction and windage losses. Second, the reactive, or magnetizing, current required by the magnetic field of the isolated alternator is provided by the excitation capacitor C, shown in parallel with the load impedance Z_L in Fig. 7. The assumptions inherent in the formulation of this equivalent circuit are made for the following analysis.

Basic Relationships

Induced Voltage and Current

If the air-gap flux varies sinusoidally, the induced voltage is[12]

$$E_1 = 4.44 k_p k_d f N_1 \Phi_{mg} \quad (43)$$

where

E_1 = the rms stator induced voltage per phase
k_p = the stator-winding pitch factor
k_d = the stator-winding distribution factor
N_1 = the number of stator series turns per phase
Φ_{mg} = the maximum value of the total air-gap flux per pole

If the expression for the total stator core loss, equation 60, is to be useful in a general solution (that is, contain only products of the variables), the assumption of a constant ratio of tooth flux density to yoke flux density is necessary. This assumption is compatible with the assumption of geometric similitude only if the number of poles is assumed constant. For simplicity of analysis, the ratio of tooth-to-yoke flux density is assumed equal to unity and both are designated as the core flux density. The total flux per pole can be expressed in terms of the air-gap flux density or the stator-tooth density:[12]

$$\Phi_{mg} = \frac{2DlB_{mg}}{p} \quad (44)$$

and

$$\Phi_{mg} = \frac{2k_l k_r \tau l B_{ms}}{p} \quad (45)$$

whence

$$B_{mg} = k_l k_r B_{ms} \quad (46)$$

where

B_{mg} = the maximum value of the air-gap flux density
B_{ms} = the maximum value of the core flux density
D = the air-gap diameter
l = the gross axial length of the core
p = the number of poles
k_l = the core stacking factor
k_r = the ratio of average stator tooth width to air-gap tooth pitch

If Φ_{mg} from equation 45 is substituted into equation 43,

$$E_1 = \frac{K_s D l_o}{p} f N_1 B_{ms} a^2 \quad (47)$$

where

$K_s \equiv 8.88 k_p k_d k_r k_l$

The rms stator current I_1 is

$$I_1 = \frac{k_{w1} A_{wo1}}{q}\frac{J_1 a^2}{N_1} \quad (48)$$

where

k_{w1} = the stator-winding space factor
A_{w1} = one half the total cross-sectional area of all the stator slots
q = the number of phases
J_1 = the rms stator current density

Impedances

The base impedance per phase Z_{b1} is

$$Z_{b1} \equiv \frac{E_1}{I_1} = \frac{qK_s D_o l_o}{p k_{w1} A_{wo1}}\frac{f N_1^2 B_{ms}}{J_1} \quad (49)$$

The stator winding resistance per phase R_1 is

$$R_1 = \frac{q\rho_{s1} l_{wo1} N_1^2}{k_{w1} A_{wo1} a} \quad (50)$$

whence

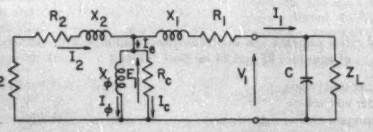


Fig. 7. Single-phase induction-alternator equivalent circuit

$$R_{ps1} \equiv \frac{R_1}{Z_{b1}} = \frac{\rho_{ps1} l_{ws1}}{K_s D_{sb}} \frac{J_1}{f B_{ms} a} \tag{51}$$

where

ρ_{s1} = the effective stator-winding resistivity

l_{ws1} = the mean length of stator-winding turn

Inspection of the equations for the various components of the leakage reactance[13] shows that the method of representing the transformer leakage inductance, equation 15, also can be applied to the alternator leakage inductance.

Thus

$$L_1 = q K_{ps1} N_1^2 \tag{52}$$

where L_1 is the stator leakage inductance per phase and K_{ps1} is defined as the stator leakage permeance factor. From equation 52, the stator leakage reactance is

$$X_1 = 2\pi q K_{ps1} f N_1^2 a \tag{53}$$

whence

$$X_{ps1} \equiv \frac{X_1}{Z_{b1}} = \frac{2\pi p k_{ws1} A_{ws1} K_{ps1}}{K_s D_{sb}} \frac{J_1 a}{B_{ms}} \tag{54}$$

Volt-Ampere Rating

The total induced volt-amperes are

$$(VA)_t \equiv q E_1 I_1 = \frac{K_s k_{ws1} A_{ws1} D_{sb}}{p} f J_1 B_{ms} a \tag{55}$$

and the output volt-amperes can be written as

$$(VA)_t \equiv q V_1 I_1 = \frac{(VA)_t}{|1/\underline{\theta} + R_{ps1} + jX_{ps1}|} \tag{56}$$

where V_1 is the terminal voltage and θ is the power-factor angle of the parallel combination of the excitation capacitors and the load. (As defined, the angle θ is negative for the induction alternator.)

Losses and Heat Transfer

The treatment of heat transfer in rotating machinery is simplified by the following assumptions: 1. The heat transferred between the stator winding and the stator core is assumed negligible. 2. The heat generated in the stator core is conducted to a heat sink. 3. The heat generated in the stator winding is conducted to the end turns and then dissipated by convection. 4. The rotor-winding and rotor-core temperatures can be higher than the corresponding stator temperatures owing to the absence of electrical insulation on the rotor. In order to obtain a general solution, the surface temperature of the stator-winding end turns is used as the thermal criterion; for the stator core, the maximum internal temperature is used.

The total stator winding loss P_{ws1} is

$$P_{ws1} \equiv q I_1^2 R_{ps1} = \rho_{s1} k_{ws1} l_{ws1} A_{ws1} J_1^2 a^2 \tag{57}$$

The heat transferred from the stator end turns is approximated by

$$P_{ws1} = h_{ws1} A_{sswa} a^2 \Delta t_{sws1} \tag{58}$$

where

h_{ws1} = the average heat-transfer coefficient for the stator end turns

A_{sswa} = the total exposed surface area (heat-transfer area) of the stator end turns

Δt_{sws1} = the average surface temperature rise of the stator end turns above the ambient or cooling-air temperature

From equations 57 and 58,

$$\Delta t_{sws1} = \frac{\rho_{s1} k_{ws1} l_{ws1} A_{ws1}}{h_{ws1} A_{sswa}} J_1^2 a \tag{59}$$

The total stator core loss P_{cs} is

$$P_{cs} = k_f D_{sb} A_{cs1} f^m B_{ms}^n a^2 \tag{60}$$

where A_{cs1} is the total cross-sectional area of the stator core perpendicular to the shaft. The heat conducted away from the core is

$$P_{cs} = K_{tscs1} a \Delta t_{ms} \tag{61}$$

where K_{tscs1} is a constant describing the model (see the discussion preceding equation 30) and Δt_{ms} is the difference between the maximum internal stator-core temperature and the heat-sink temperature. From equations 60 and 61,

$$\Delta t_{ms} = \frac{k_f D_{sb} A_{cs1}}{K_{tscs1}} f^m B_{ms}^n a^2 \tag{62}$$

Excitation Requirement

If a saturation factor k_s is defined as the ratio of the total magnetomotive force (mmf) required by the entire magnetic circuit to the mmf required by the air gap and if equation 46 for B_{ms} is used, the total magnetizing current I_ϕ is

$$I_\phi = \frac{p K_\phi k_s g_o B_{ms} a}{q N_1} \tag{63}$$

where $K_\phi = \pi k_f k_r / 2\sqrt{2} \mu_o k_p k_d$ and g is the effective air-gap length. The magnetizing volt-amperes are, from equations 47 and 63,

$$(VA)_\phi \equiv q E_1 I_\phi = K_s K_\phi k_s D_{sb} l_{cs} g_o f B_{ms}^2 a^2 \tag{64}$$

Because the energy stored in the leakage fields is very small compared with the energy stored in the mutual air-gap field, the capacitance C per phase of the excitation capacitors is

$$C = \frac{I_\phi}{2\pi f E_1} = \frac{p^3 K_\phi k_s g_o}{2\pi q K_s D_{sb}} \frac{1}{f^3 N_1^3 a} \tag{65}$$

If the volume of an excitation capacitor is approximately proportional to the product of the capacitance times the capacitor voltage rating, the total volume V_c of the q capacitors is

$$V_c \propto q E_1 C = \frac{p K_\phi k_s g_o}{2\pi} \frac{B_{ms} a}{f N_1} \tag{66}$$

The Generalized Prototype

The parameters that are maintained constant include the output voltage and current, the stator-winding temperature, the number of poles, the flux density, and the ambient and the heat-sink temperatures. The variables a, J_1, and N_1 can be found from the simultaneous solution of equations 47, 55, and 59 for E_1, $(VA)_t$, and Δt_{sws1}.

The size factor is

$$a = \left[\frac{\rho (VA)_t (\rho_{s1} k_{ws1} l_{ws1} A_{ws1})^{1/2}}{(K_s k_{ws1} A_{ws1} D_{sb} l_{cs} B_{ms})^{1/2} (h_{ws1} A_{sswa} \Delta t_{sws1})^{1/2}} \right]^{2/7} f^{-2/7} \tag{67}$$

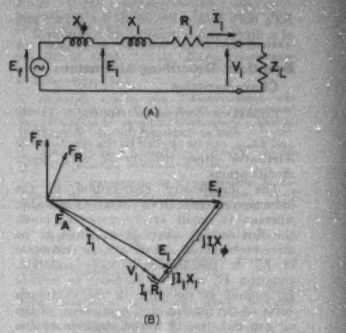


Fig. 8. Single-phase equivalent circuit and vector diagram for synchronous alternators

The weight of the electromagnetic portions of the alternator is proportional to the cube of a:

$$W_t \propto \left[\frac{\rho (VA)_t}{\Delta t_{sws1}^{1/3}} \right]^{3/7} f^{-3/7} \tag{68}$$

The relationships between the alternator variables and frequency for constant $(VA)_t$ are summarized in Table II. Only two exponents are functions of the core-material characteristics:

$$P_{cs} \propto f^{(\alpha - 3\beta)} \tag{69}$$

and

$$\Delta t_{ms} \propto f^{(\alpha - 3\beta)} \tag{70}$$

From equations 56 and 68,

$$W_t \propto \left[\frac{\rho (VA)_t}{\Delta t_{sws1}^{1/3}} |1/\underline{\theta} + R_{ps1} + jX_{ps1}| \right]^{3/7} f^{-3/7} \tag{71}$$

In order to compute curve *3* of Fig. 2, the impedance terms in equation 71, which are functions of frequency, have been determined from the exponents given in Table II and from measurements on a 3-phase 400-cps 500-watt 115-volt 24,000-rpm induction alternator; at 400 cps, $R_{ps1} = 0.0629$ and $X_{ps1} = 0.0755$.

Appendix III. Wound-Field Synchronous Alternators

A balanced polyphase cylindrical-rotor alternator supplying a balanced load can be represented by the equivalent circuit of Fig. 8(A) where E_f is the excitation voltage produced by the field mmf F_F in the absence of armature reaction, E_1 is the air-gap voltage produced by the resultant mmf F_R, and X_ϕ is the magnetizing reactance (or the reactance of armature reaction) and represents the effect of the armature-reaction mmf F_A. The corresponding vector diagram is shown in Fig. 8(B). For a salient-pole alternator, the reluctances of the direct and the quadrature axes of the rotor are unequal, and the voltage drop represented in Figs. 8(A) and 8(B) by

I_1X_ϕ may be considered as the vector sum of a direct and a quadrature component.

Equations Describing Alternator Characteristics

Equations derived in Appendix II for E_1, I_1, R_{pu1}, X_{pu1}, $(VA)_t$, $(VA)_t$, Δt_{sw1}, and Δt_{m1}, can be applied to the synchronous alternator either directly or with minor modifications.

The flux-density distribution in the induction machine is very nearly sinusoidal, whereas in small synchronous machines, the flux-density distribution tends to be more rectangular. Therefore, the definition of K_s in the induced-voltage equation, equation 47, must be modified to include k_b, defined as the ratio of the maximum value of the fundamental component of the air-gap flux density to the maximum value of the actual air-gap flux density; thus, $K_s = 8.88 k_p k_d k_i k_b k_r$.

Inspection of equations for the leakage and the magnetizing reactances of synchronous alternators[13,14] reveals that, as long as geometric similitude is maintained, these reactance components and the corresponding inductance terms can be represented by equations 53 and 52, providing the proper definition of K_{pq} is used. Therefore, the various per unit reactances are all proportional to X_{pu1}, given by equation 54; thus

$$X_{puk} = \frac{2\pi p k_{w1} A_{w01} K_{pook}}{K_s D_a l_a} \frac{J_1 a}{B_{mi}}, \qquad k = 1, \phi, \phi d, \phi q \quad (72)$$

where $X_{\phi d}$ and $X_{\phi q}$ are the direct-axis and quadrature-axis magnetizing reactances.

Field Winding

The mmf per pole that must be supplied by the field winding is equal to the vector difference between the resultant mmf per pole and the armature-reaction mmf per pole, as illustrated in Fig. 8. The armature reaction mmf can be written as

$$F_A = \frac{2\sqrt{2} q k_p k_d N_1 I_1}{\pi p} \equiv K_A N_1 \quad (73)$$

If the space harmonics of the resultant air-gap flux-density distribution are neglected, the mmf required to establish the resultant flux is

$$F_R = \frac{k_b k_{\phi L o}}{\sqrt{2}\mu_o} B_{mg} a \quad (74)$$

Substitution for B_{mg} from equation 46 yields

$$F_R = \frac{k_b k_{\phi L} k_r k_{\phi L o}}{\sqrt{2}\mu_o} B_{mi} a \equiv K_{Rao} B_{mi} a \quad (75)$$

The magnitude of the mmf that must be supplied by the field is

$$F_F = \left| \frac{F_R}{\psi_R} - \frac{F_A}{\psi_A} \right| = \left| K_{Rao} B_{mi} a \underline{/\psi_R} - K_A N_1 \underline{/\psi_A} \right| \quad (76)$$

where ψ_R and ψ_A are the phase angles of F_R and F_A, respectively. Because the relative magnitudes of K_{Rao} and K_A are not known in general, the summation form of equation 76 limits its usefulness in a general solution. Therefore, the following approximation for F_F is assumed:

$$F_F \approx K_{Fao} B_{mi} a \quad (77)$$

The validity of this approximation will be verified later. The required field current I_2 is

$$I_2 = \frac{F_F}{K_{w2} N_2} = \frac{K_{Fao} B_{mi} a}{K_{w2} N_2} \quad (78)$$

where K_{w2} is the product of the field-winding pitch and distribution factors and N_2 is the number of field turns per pole.

The total field-winding resistance R_2 can be written

$$R_2 = \frac{p^2 \rho_{e2} l_{w02} N_2^2}{k_{w2} A_{w02} a} \quad (79)$$

where

ρ_{e2} = the effective resistivity of the field conductor material at the average winding temperature

l_{w2} = the mean length of field-winding turn

A_{w2} = one half the gross cross-sectional area of the field winding

k_{w2} = the field-winding space factor

The total power P_{w2} dissipated in the field winding is

$$P_{w2} = I_2^2 R_2 = \frac{p^3 K_{Fao}^2 \rho_{e2} l_{w02} B_{mi}^2 a}{k_{w2} K_{w2}^2 A_{w02}} = K_{Lao} B_{mi}^2 a \quad (80)$$

Cooling of the rotor by convection is assumed. If heat transfer between winding and core is assumed negligible, the winding loss also can be written as

$$P_{w2} = h_{w2} A_{sw02} a^2 \Delta t_{sw2} \quad (81)$$

where

h_{w2} = the average heat-transfer coefficient of the field winding

A_{sw2} = the effective heat-transfer area of the field winding

Δt_{sw2} = the average surface temperature rise of the winding above the cooling-air temperature

From equations 80 and 81,

$$\Delta t_{sw2} = \frac{K_{Lao}}{h_{w2} A_{sw02}} \frac{B_{mi}^2}{a} \quad (82)$$

The Generalized Prototype

As explained in the text, the temperature rises of the 400-cps model are assumed to be

$$\Delta t_{sw01} = \Delta T_m$$

$$\Delta t_{sw02} = \Delta T_m$$

$$\Delta t_{m1a1} = (1/2)\Delta T_m$$

where ΔT_m is the maximum allowable temperature rise. As the frequency is increased from 400 cps, t_{sw1} and t_{sw2} are assumed constant; t_{m1} increases with frequency and reaches the maximum value at 1,500 cps. For frequency increases above this value, t_{sw1} and t_{m1} are maintained constant; t_{sw2} decreases.

Solution for Constant t_{sw1} and t_{sw2}; Case I

If $(VA)_t$, Δt_{sw1}, and Δt_{sw2}, given by equations 55, 59, and 82 are assumed constant, the size factor and the weight are

$$a = \left[\frac{p(VA)_t (\rho_{e1} k_{w1} l_{w01} A_{w01})^{1/3} K_{Lao}^{1/3}}{(K_s k_{w1} A_{w01} D_a l_a)(h_{w1} A_{sw01} \times \Delta t_{sw1})^{1/3} (h_{w2} A_{sw02} \Delta t_{sw2})^{1/3}} \right]^{1/4} f^{-1/4} \quad (83)$$

and

$$W_t \propto \left[\frac{p(VA)_t}{(\Delta t_{sw1} \Delta t_{sw2})^{1/3}} \times \left| 1/\underline{\theta} + R_{pu1} + jX_{pu1} \right| \right]^{3/4} f^{-3/4} \quad (84)$$

Solution for Constant t_{sw1} and t_{m1}; Case II

If t_{sw1} and t_{m1} are assumed constant, the size factor and weight are given as

$$a = \left[\frac{p(VA)_t (\rho_{e1} k_{w1} l_{w01} A_{w01})^{1/3}}{(\gamma k_d l_a A_{w1})^{1/3}} \times \frac{2\delta}{7\beta - 4} \right] \frac{2\delta}{(K_s k_{w1} A_{w01} D_a l_a)(h_{w1} A_{sw01} \times \Delta t_{sw1})^{1/3} (K_{Lo1} \Delta t_{m1})^{1/3}} \Bigg]^{\frac{2\delta}{7\beta - 4}} f^{-\frac{2}{3}\left(\frac{\beta - \alpha}{7\beta - 4}\right)} \quad (85)$$

and

$$W_t \propto \left[\frac{\gamma p (VA)_t}{\Delta t_{sw1}^{1/3} \Delta t_{m1}^{1/3}} \times \left| 1/\underline{\theta} + R_{pu1} + jX_{pu1} \right| \right]^{\frac{6\delta}{7\beta - 4}} f^{-\frac{6}{7}\left(\frac{\beta - \alpha}{7\beta - 4}\right)} \quad (86)$$

The relationships between other alternator characteristics and frequency are given in Table II. Curve *1* of Fig. 2 is plotted from equations 84 and 86. The same values for R_{pu1} and X_{pu1} at 400 cps used in Appendix II also are used here.

Validity of the Approximation for F_F

The validity of the approximation for F_F, equation 77, now can be established by substituting the appropriate functions of frequency from Table II into equations for F_A, F_R, and F_F. For constant t_{sw1} and t_{sw2}, this substitution shows that the magnitudes of the two vector components of F_F vary with frequency in the same manner; that is,

$$F_A \propto F_R \propto f^{-3/8} \quad (87)$$

Furthermore, because $R_{pu1} \ll 1$ and the synchronous reactance $(X_\phi + X_1)$ is invariant with frequency (see Table II), the angle between these two mmf components is invariant with frequency. Therefore, the F_F-frequency relationships represented by equations 76 and 77 are the same, and equation 82 is a valid expression for the field-winding temperature.

Similar substitutions for case II shows that the mmf given by equation 77 and also the field-winding temperature rise, given by equation 82, are unrealistically low. On the other hand, if the mmf were approximated as $F_F \propto N_1$, t_{sw2} is independent of frequency. However, because the field-winding temperature is not a criterion for the determination of the generalized prototype for case II, the uncertain relationship between F_F and f affects only the determination of Δt_{sw2} which is between ΔT_m and $0.30\Delta T_m$ at 5,000 cps.

Appendix IV. Permanent-Magnet Alternators

The primary difference between the permanent-magnet alternator and the wound-field alternator can be characterized by the relationships between the mmf and flux for the respective rotors; in the absence of armature reaction, the rotor mmf of a given permanent-magnet alternator is a function of the magnet stabilization and the flux in the magnet; the field mmf of a given wound-field machine is dependent only on the field current. In both machines, however, the air-gap flux under load conditions is determined by both the rotor mmf and the armature-reaction mmf. Therefore, the equivalent circuit of Fig. 8(A) can be used to represent the permanent-magnet alternator if X_ϕ is related to the average slope of the appropriate minor hysteresis loop of the magnet.[9]

Because the stator core temperature of the prototype is assumed equal to the maximum permissible value, the core flux density must be decreased with increasing frequency. If geometric similitude of the magnet is maintained and if the method of stabilizing the magnet is not altered, it can be shown that the flux density in the magnet and in the stator core of the prototype is equal to or greater than that in the model. Therefore, the flux density in the stator must be artificially decreased with increasing frequency to prevent excessive core temperatures; the characteristics of the magnet do not influence the scaling process. Accordingly, the analysis of the permanent-magnet alternator is virtually identical with that of the wound-field alternator with constant i_{fnl} and i_{mn}; accordingly, the expressions for the size factor and weight, given by equations 85 and 86, and for other alternator characteristics, given in Table II, are applicable to the permanent-magnet alternator. The weight-frequency characteristics in Fig. 2 are computed with the same values of R_{pu} and X_{pu} used in Appendix II.

The decrease in flux density necessary to maintain constant stator core temperature can be realized either by stabilizing the prototype magnet to a value of flux density lower than that employed for the model or by altering the magnet dimensions. The latter method would provide a lighter alternator.

References

1. Determination of an Optimum Primary Power Frequency and Voltage for Missiles, S. K. Grinnell, R. E. Turkington. *Proceedings*, Symposium on Auxiliary Power for Guided Missiles, Office of the Assistant Secretary of Defense for Research and Development, Washington, D. C., vol. III, Feb. 1957, pp. 22–44 (confidential).

2. Determination of an Optimum Primary Power Frequency and Voltage for Missiles, R. W. Mann, S. K. Grinnell, R. E. Turkington. *Report no. 107*, Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Nov. 12, 1957 (confidential).

3. Study of Aircraft Alternating-Current Generating and Voltage-Regulating Systems, R. M. Saunders, D. C. White. *WADC Technical Report 53-311*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Aug. 1955, chaps. 5 and 6.

4. Design of a High-Speed Induction Generator, W. H. Gable. *S. M. Thesis*, Department of Electrical Engineering, Massachusetts Institute of Technology, 1953.

5. Hipersil Cores for Transformers and Reactors, Westinghouse Catalog, Greensville, Pa., Transformer Division, Westinghouse Electric Corporation, Jan. 10, 1955, sect. 44-550.

6. Armco Thin Electrical Steels, Armco Steel Corporation, Middletown, Ohio.

7. Design Principles of Flux-Switch Alternators, S. E. Rauch, L. J. Johnson. *AIEE Transactions*, vol. 74, pt. III, Dec. 1955, pp. 1261–68.

8. Fuels and Prime Movers for Rotating Auxiliary Power Units, R. W. Mann. *Proceedings*, Symposium on Auxiliary Power for Guided Missiles, vol. III, pp. 95–117.

9. Components Handbook (book), J. F. Blackburn, Radiation Laboratory Series, McGraw-Hill Book Company, Inc., New York, N. Y., vol. 17, 1949, pp. 117–18.

10. Transformer Engineering (book), L. F. Blume, A. Boyajian, G. Camilli, T. C. Lennox, S. Minneci, V. M. Montsinger, John Wiley & Sons, Inc., New York, N. Y., second edition, 1951.

11. Review of Industrial Applications of Heat Transfer to Electronics, J. Kaye. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, Aug. 1956, pp. 977–91.

12. The Nature of Polyphase Induction Machines (book), P. L. Alger. John Wiley & Sons, Inc., 1951.

13. The Calculation of the Armature Reactance of Synchronous Machines, P. L. Alger. *AIEE Transactions*, vol. 47, Apr. 1928, pp. 493–513.

14. Calculation of Synchronous Machine Constants—Reactances and Time Constants Affecting Transient Characteristics, L. A. Kilgore. *Ibid.*, vol. 50, Dec. 1931, pp. 1201–14.

Discussion

D. J. Hanrahan (U. S. Naval Research Laboratory, Washington, D. C.): Fig. 2 shows relative weight as a function of frequency for wound-field, permanent-magnet, and induction alternators. It appears that the least weight reduction is obtained with the permanent-magnet generator. The slopes of these curves differ because of three different thermal criteria. It is assumed that the model permanent-magnet machine is designed for maximum temperature rise in stator winding and iron, and that these temperatures are to be held while the frequency (speed) is increased.

This assumption, and the resulting slope, are incorrect. The design of the model permanent-magnet machine is excitation, not temperature, limited. The saturation density of permanent-magnet material is less than that of armature steel so that tooth densities tend to be low. To saturate the yoke would increase the effective air gap which is already critical. Consequently, the stator iron temperature rise of a properly designed permanent-magnet generator at 400 cycles is necessarily below the thermal limit.

It would seem to follow that the same criterion as for the induction alternator can be used, giving a greater decrease in weight with increase in frequency. However, it is not clear that a simple scaling process on the foregoing thermal basis would yield the optimum design at each frequency. Because the permanent-magnet machine is excitation limited, we believe that scaling must be on a magnetic basis if maximum kilovolt-amperes per pound is to be obtained.

The paper states that the equivalent circuit of Fig. 8(A) can be used to represent the permanent-magnet alternator if X_ϕ is related to the average slope of the appropriate minor hysteresis loop of the magnet. If the source voltage (E_f) in Fig. 8(A) is that generated by the magnet mmf acting alone, X_ϕ is the conventional reactance of armature reaction, simply related to the air-gap permeance,[1] and E_f increases with demagnetizing load as implied in the paper.

But presumably E_f is intended to represent the no-load voltage; the foregoing reactance is then reduced by a factor involving air-gap and pole-leakage permeances as well as the minor loop slope.[2] Either circuit will account for the self-regulating tendency of the permanent-magnet generator.

References

1. Permanent Magnet Generators, Part I—Theory, D. J. Hanrahan, D. S. Toffolo. *AIEE Transactions*, vol. 76, pt. III, Dec. 1957, pp. 1098–1103.

2. Synchronous Machines with Rotating Permanent-Magnet Fields, Part II—Magnetic and Electrical Design Considerations, Fritz Strauss. *Ibid.*, vol. 71, pt. III, Oct. 1952, pp. 887–93.

R. E. Turkington: The voltage source in the equivalent circuit for the permanent-magnet alternator (see Appendix IV) is the no-load voltage. The reactance X_ϕ is a function of the air-gap, leakage-path, and magnet permeances[1] and the minor-loop slope, as indicated in the paper. Although either equivalent circuit mentioned by Mr. Hanrahan could be used, the one selected facilitates an understanding of the scaling process and of the results.

Table IV. Permanent-Magnet Characteristics as Functions of Frequency for Constant Induced Volt-Amperes

Characteristics, M(f) $\propto f^{\text{exp}}$	Exponent	M(4,000) / M(400)
Weight, W_t	−1	0.100
Size factor, σ	−1/3	0.465
Flux density, B_{mt}	*	1.000
Stator current density, J_s	1/3	2.15
Number of stator turns, N_t	−1/3	0.465
Per-unit resistance, R_{pu}	−1/3	0.465
Per-unit reactance, X_{pu}	0	1.000
Stator conductor area, A_{st}	−1/3	0.465
Stator winding loss, P_{wt}	−1/3	0.465
Stator core loss, P_{it}	0.25	1.772
Stator core temperature, $\Delta\theta_{sti}$	0.58	3.80
Stator winding temperature, $\Delta\theta_{wt}$	1/3	2.15
Demagnetizing ampere-turns, H_d	*	1.000

* Full-load flux density and armature demagnetizing mmf maintained constant.

Mr. Hanrahan's comments with regard to the assumption of constant stator winding and iron temperatures are a valuable criticism. H. Cooke of the TKM Electric Corporation has made the same comment and has suggested that the temperature rises of the stator winding and of the core of a typical 400-cps permanent-magnet alternator are on the order of 50 C, well below the allowable maxima.

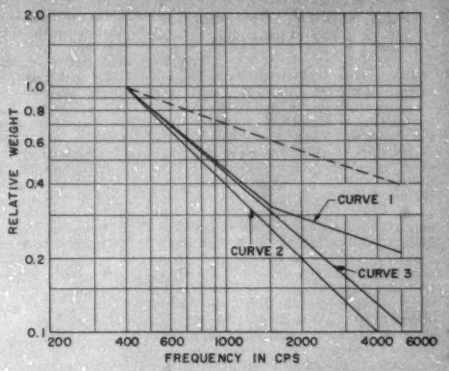
To illustrate a more realistic functional relationship between frequency and the various alternator characteristics, an analysis has been made assuming the full-load flux density B_{mt} and the full-load demagnetizing armature mmf H_d are constant. The results are shown in Table IV and in Fig. 9. Also shown in the figure for comparison are the other weight-frequency curves shown originally in Fig. 2 of the paper. These results show a very substantial decrease in weight with increasing frequency. Furthermore, the table shows that the winding and iron temperature rises increase significantly with increasing frequency. However, owing to the low values of temperature rise in the 400-cps model, the temperature rise in the winding is 115 C and the core is 216 C at 5,000 cps, values still below the allowable maxima for short-duty-cycle operation.

The full-load flux density and direct-axis demagnetizing ampere-turns are assumed constant in view of the following reasoning. If the rotor is air-stabilized, the intersections of the magnet B-H curve and the out-stator permeance line[1] and of the B-H curve and the air-gap permeance line are not functions of the size factor a. Thus, as alternator size is varied while maintaining geometric similitude, the values of the stabilized flux density (i.e., the value of flux density that exists when the rotor is removed from the stator) and of the no-load flux density are constant. Therefore, movement of the full-load operating point in the B-H plane is limited to movement along the invariant minor loop. This fact, together with the realization that the full-load operating point of the model alternator would probably be near the maximum-energy point of the magnet, indicates that the assumptions of constant B_{mt} and H_d are realistic. This reasoning also can be applied to short-circuit-stabilized alternators if the armature resistance is small when compared with the synchronous reactance.

I am grateful to Mr. Hanrahan for pointing out this error and for this opportunity to present the more realistic results.

Fig. 9. Relative weight of alternators as a function of frequency

Curve 1—Wound-field synchronous alternator
Curve 2—Permanent-magnet alternator
Curve 3—Induction alternator
The dashed curve shows curve 2 of Fig. 2



Reference

1. See reference 2 of Mr. Hanrahan's discussion.

Properties of Electrical Insulation at Ultrahigh Temperatures

G. I. DUNCAN
ASSOCIATE MEMBER AIEE

M. M. FELGER
NONMEMBER AIEE

THE ERA of high-speed nuclear-propelled aircraft and missiles has brought with it the need for electronic equipment which is both temperature and radiation tolerant. Present temperature and radiation goals as established by the U. S. Air Force for electronic power transformers are 500 C (degrees centigrade) ambient temperature and intense nuclear radiation. These are quite different from the 125 C ambient temperature and no radiation requirements in today's electronic equipment. Major technological advances will have to be made in all phases of electronic design and manufacture to realize these radical environmental changes.

The Air Force, recognizing the need for temperature- and radiation-tolerant electronic equipment, has initiated various contracts for developing this equipment. The work described in this paper was done as part of an Air Force Contract to develop temperature- and radiation-tolerant electronic power transformers.

Materials ultimately set the limits on how high a temperature and radiation intensity the components and equipment can be designed for and still perform their assigned function. For electronic transformers, three essential materials are coil insulation (sheet and spool), insulated wire, and the magnetic iron core. This paper presents the results of an investigation to determine the effects of 500 C temperature on sheet and spool insulations for electronic transformers. The effects of radiation on these materials was not a part of this investigation.

Table I. Materials Selected for Evaluation

Material	Thickness, Mils
Sheet Insulation	
Aluminum silicate paper	10 and 20
Asbestos paper, silicone impregnated	3
Glass paper	4
Integrated phlogopite mica, silicone bonded	2–3
Quartz fiber cloth	4
Recombined muscovite mica paper, silicone bonded	4
Recombined muscovite mica paper—glass cloth, silicone bonded	4
Spools	
Recombined muscovite mica paper, aluminum phosphate bonded	31–35
Recombined muscovite mica paper, silicone bonded	23–28
Steatite	57

Paper 58-115, recommended by the AIEE Electrical Insulation Committee, formerly the Dielectrics Committee of the Communication and Electronics Division, and the Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted October 14, 1957; made available for printing November 13, 1957.

G. I. Duncan and M. M. Felger are with the General Electric Company, Fort Wayne, Ind.

This work was done as part of U. S. Air Force Contract No. AF33(616)-3623.

Table II. Initial Physical Properties of Ultrahigh-Temperature Layer Insulation

Layer Insulation	Thickness, Mils	Tensile Strength, Pounds/Inch Width	Elongation, %	Flexibility 1/8 Inch Mandrel, 180° Bend
Aluminum silicate paper, source A	17.3	0.32	0.76	satisfactory
Aluminum silicate paper, source B	12.2	0.30	0.7	satisfactory
Asbestos paper, silicone impregnated	3.0	9.5	1.3	satisfactory
Glass paper	3.3	2.4	0.55	satisfactory
Integrated phlogopite mica, silicone bonded	2.7	11.2	0.25	satisfactory
Quartz fiber cloth	3.5	2.2	1.1	satisfactory
Recombined muscovite mica paper, silicone bonded	4.1	11.1	0.42	satisfactory
Recombined muscovite mica paper-glass cloth, silicone bonded	5.3	27.1	2.3	satisfactory

Table III. Physical Condition of Ultrahigh-Temperature Insulation After 1,000 Hours at 600 C

Material	Condition
Layer Insulation	
Aluminum silicate paper, 10 mils	Little change
Asbestos paper, silicone impregnated	Complete failure by physical disintegration
Glass paper	Complete failure by physical disintegration
Integrated phlogopite mica, silicone bonded	Little change, except decrease in flexibility
Recombined muscovite mica paper, silicone bonded	Considerable loss of flexibility; 50% increase thickness
Recombined muscovite mica paper-glass cloth, silicone bonded	Glass cloth disintegrated; complete failure
Spool Material	
Recombined muscovite mica paper, aluminum phosphate bonded	Increased in thickness 40%; otherwise good condition
Recombined muscovite mica paper, silicone bonded	Brittle and delaminated
Steatite	No observable change
Spools	
Recombined muscovite mica paper, aluminum phosphate bonded	Good condition
Recombined muscovite mica paper, silicone bonded	Badly delaminated
Steatite	Good condition

Selection of Materials and Test Criteria

The materials listed in Table I were selected for evaluation. Since the sheet insulations would have to withstand winding into coils, the following initial physical properties were measured: 1. elongation, 2. flexibility, and 3. tensile strength.

The specified thermal life of the coil insulations was 1,000 hours in a maximum ambient temperature of 500 C. It was estimated that the temperature rise of the transformers in which the insulation would be used would be about 100 C. This would bring the total temperature up to 600 C. To determine the effects of the ultra-high temperatures on the coil insulations, electrical properties were measured both at temperature and after heat aging. The test temperatures selected were 300 C, 400 C, 500 C, and 600 C. One thousand hours and 600 C were selected for the heat aging conditions. The electrical properties picked to be measured were 1. electric strength, 2. volume resistivity, 3. dissipation factor at 100 and 1,000 cps (cycles per second), and 4. dielectric constant at 100 and 1,000 cps.

Physical Properties

Elongation and tensile strength were measured on a standard constant speed crosshead tensile machine. Flexibility was determined by a 180-degree bend over 1/8-inch mandrel.

Table II gives the initial physical properties of layer insulation. All of these materials have relatively low tensile strength and elongation, but the better ones can be fabricated into coils with proper handling.

Table III gives the condition of all materials after 1,000 hours at 600 C. The silicone-bonded recombined mica had

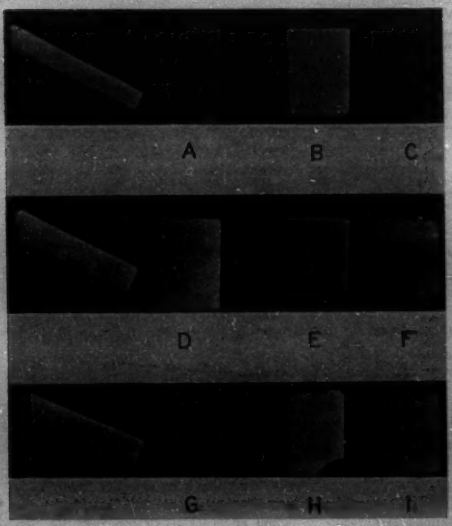


Fig. 1. Physical condition after 1,000 hours at 600 C

A—Steatite
B—Aluminum silicate paper
C—Glass paper
D—Silicone-bonded recombined mica paper
E—Silicone-bonded recombined mica paper
F—Silicone-bonded recombined mica paper-glass cloth
G—Aluminum phosphate bonded recombined mica paper
H—Silicone impregnated asbestos paper
I—Silicone-bonded integrated phlogopite mica

Fig. 2 (below). Electric strength electrode assembly

Fig. 3 (left). Electrode assembly for measurement of electric strength and insulation resistance

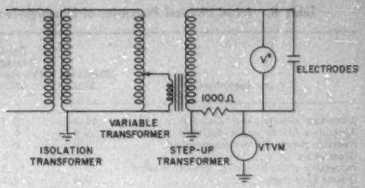


Fig. 4 (above right). Circuit for electric strength measurements

*<5,000 volts: rectifier-type voltmeter
>5,000 volts: electrostatic voltmeter

the best thermal resistance. Fig. 1 shows the physical condition of all materials after aging 1,000 hours at 600 C.

Electric Strength

The initial electric strength of materials at room conditions have little value for materials used at ultra-high temperatures. Therefore this property was determined after various thermal treatments at the treatment temperatures. One condition was an initial aging of 40 hours to drive off any volatile or easily decomposable materials.

The initial dielectric strength at room temperature for flat insulation was measured according to American Society for Testing Materials Method *D149*. The room temperature electric strength of spools was measured between an inner electrode of tightly packed steel wool and an outer electrode of lead foil.

At elevated temperatures the electrodes and samples were mounted on a sliding frame with a door of high temperature asbestos board replacing the regular door of the furnace. The electrode connections passed through holes in the door without touching and were connected by flexible leads to terminals in a second asbestos board mounted 6 inches from the door. Figs. 2 and 3 show the electrode assembly from the rear and partially removed from the furnace. This arrangement prevented current leakage through the door at elevated temperatures. The electrodes were 2 inches in diameter and of standard shape made from *304* stainless steel. Stainless steel leads with a right-angle bend came through the false door openings.

At elevated temperature the electric strength measurements of spools were made in the furnace on the sliding frame. Stainless steel wool was the inner electrode. The outer electrode was a 1-inch wide conducting ceramic band.

Fig. 4 shows the circuit used for measuring electric strength at elevated temperatures. Failure was indicated by a sudden increase of current on the vacuum tube voltmeter.

However, on some materials only a limiting value could be obtained because of small sample size or increased conductance at the elevated temperatures. The materials that showed high conductance were silicone bonded recombined muscovite mica, aluminum phosphate bonded recombined muscovite mica paper, glass paper, and steatite.

The results are given in Table IV. The best retention of electric strength by layer insulations was obtained with silicone-bonded integrated phlogopite mica and silicone-bonded recombined muscovite mica paper. Three materials, silicone-impregnated asbestos paper, glass paper, and recombined muscovite mica paper-glass, cloth, silicone bonded, failed physically. The best spool material was aluminum phosphate-bonded recombined muscovite mica, even though it increased in thickness during aging.

Table IV. Electric Strength of Ultrahigh-Temperature Insulation

Material	Nominal Thickness, Mils	Initial Electric Strength, Volts/Mil	Electric Strength, Volt/Mil, After Initial Aging: 40 Hrs. at 300 C	40 Hrs. at 400 C	40 Hrs. at 500 C	40 Hrs. at 600 C	Electric Strength, Volt/Mil, After 1,000 Hrs. at 600 C
Layer Insulation							
Integrated phlogopite mica, silicone bonded	2-3	440	260	280	260	280	266
Aluminum silicate paper (A)	10	233	114	106	97	112	105
Aluminum silicate paper (B)	20	176	96	83	86	88	86
Asbestos paper, silicone impregnated	8	410	270	330	290	280	Brittle
Glass paper	4	200	175	160	145	77	Disintegrated
Recombined muscovite mica paper, silicone bonded	4	390	390	455	485	455	330
Recombined muscovite mica-glass cloth, silicone bonded	4-5	350	260	285	270	240	Glass brittle
Quartz fiber cloth	4	147	134	128	124		
Flat Spool Material							
Recombined muscovite mica paper, silicone bonded	23-28	800	300	300	300	300	300
Recombined muscovite mica paper, aluminum phosphate bonded	31-35	700	*	*	*	50	50
Steatite	57	†	†	†	†	†	†
Spools			Total Volts				
Recombined muscovite mica paper, silicone bonded		12,000	7,900	7,600	6,000	4,800	3,700
Recombined muscovite mica paper, aluminum phosphate bonded		19,000	12,600	8,300	8,700	6,700	2,700
Steatite		25,000	12,600	6,300	3,800	2,800	3,000

* Samples too small for measurement.
† Samples not available.

Table V. Volume Resistivity of Ultrahigh-Temperature Insulations

Material	Condition: As Received, Measured at 25	40 Hrs. at 300, Measured at 300 C	40 Hrs. at 400, Measured at 400 C	40 Hrs. at 500, Measured at 500 C	40 Hrs. at 600, Measured at 600 C	1,000 Hrs. at 600, Measured at 600 C
Layer Insulation						
Aluminum silicate paper (A)	3.9×10^{13}	4.4×10^{13}	3.3×10^{11}	1.9×10^{10}	6.6×10^{9}	2.1×10^{9}
Integrated phlogopite mica, silicone bonded	3.3×10^{13}	6.0×10^{13}	8.1×10^{13}	3.3×10^{11}	7.5×10^{10}	1.0×10^{10}
Asbestos paper, silicone impregnated	1.9×10^{13}	1.7×10^{13}	3.1×10^{12}	9×10^{11}	2.0×10^{11}	*
Glass paper	3.8×10^{13}	8.0×10^{13}	9.5×10^{9}	4.1×10^{9}	6.0×10^{8}	*
Recombined muscovite mica paper, silicone bonded	1.4×10^{13}	8.2×10^{13}	3.3×10^{13}	5.4×10^{10}	1.5×10^{10}	6.5×10^{9}
Recombined muscovite mica-glass cloth, silicone bonded	3.9×10^{14}	8.7×10^{13}	3.6×10^{13}	3.0×10^{11}	8.6×10^{10}	*
Aluminum silicate paper (B)	8.3×10^{13}	8.4×10^{13}	5.3×10^{11}	3.8×10^{10}	6.8×10^{9}	1.4×10^{9}
Quartz fiber cloth	3.4×10^{8}	4.2×10^{13}	1.9×10^{11}	3.5×10^{10}	3.1×10^{9}	
Flat Spool Material						
Recombined muscovite mica paper, silicone boaded	4.2×10^{13}	1.1×10^{13}	6.7×10^{11}	6.8×10^{10}	1.5×10^{10}	5.2×10^{9}
Recombined muscovite mica paper, aluminum phosphate bonded	6.7×10^{13}	2.5×10^{13}	2.5×10^{11}	3.4×10^{10}	2.2×10^{9}	3.4×10^{8}
Steatite†	3.4×10^{13}	1.1×10^{13}	1.0×10^{11}	8.0×10^{10}	8.0×10^{10}	3.3×10^{10}

* Material in too bad a condition to measure.

† At elevated temperatures, samples were cracked by electrodes even when preheated. (Values are the best estimate obtainable.)

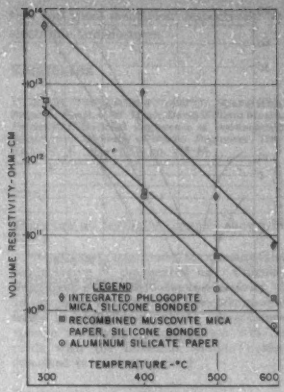


Fig. 5. Volume resistivity

Insulation Resistance

The measurements of this property were made with the same electrode and furnace assembly as described for electric strength. The electrode area was 2.4 square inches. Resistance was measured with a General Radio Megohm Bridge, type *544B*, at 100 volts direct current. These values were then calculated to volume resistivity and are given in Table V.

In all theories of conduction through insulations, it is assumed that the charge-carrying entity (electron, hole, or ion) must overcome some energy barrier in order to be able to migrate in an electric field and that the thermal activity supplies this energy to a few of the carriers not in the conducting state. These assumptions lead to the conclusion that the resistivity should increase exponentially with the reciprocal of the absolute temperature.[1]

The volume resistivities for three materials are plotted against the reciprocal of absolute temperature in Fig. 5. In general, there is a decrease in volume resistivity of one decade for each 100 C rise in temperature above 300 C.

Dissipation Factor and Dielectric Constant

Because of the difficulty of making the dissipation factor and dielectric constant measurements and the limited amount of time, determinations were made on only two materials. The best sheet insulation and the best spool insulation as determined by previous tests were selected for measurements. These materials were the silicone-bonded integrated phlogopite mica and the aluminum phosphate-bonded recombined muscovite mica.

Samples $1^3/_4$ inches $\times$ $2^1/_2$ inches were prepared from each of the selected materials. Part of the samples was heat aged for 1,000 hours at 600 C. The remainder of the samples was conditioned for 40 hours at 600 C. Gold electrodes were applied to both sides of each sample by vacuum deposition. The test specimens were then placed in a special test fixture, see Fig. 6. The dissipation factor and capacitance were measured at 100 cps with a General Radio Capacitance Bridge, type *716-C*, as the temperature was increased to 600 C. If the measured values exceeded the limit of the bridge at 100 cps, measurements were then made at 1,000 and 10,000 cps. Figs. 7(A), 7(B) and 8(A), 8(B) give plots of the dissipation factor of silicone-bonded integrated phlogopite mica and aluminum phosphate-bonded recombined muscovite mica at elevated temperatures. Figs. 9(A), 9(B) and 10(A), 10(B) give plots of the dielectric constant at elevated temperatures.

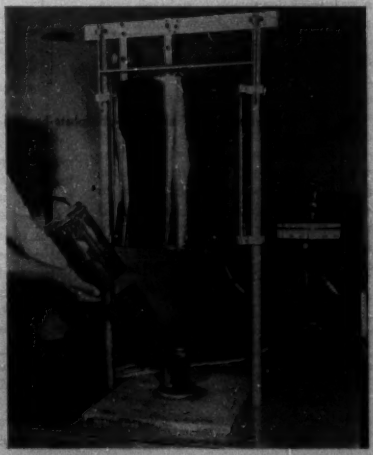


Fig. 6. Test fixture for dissipation factor and capacitance measurements

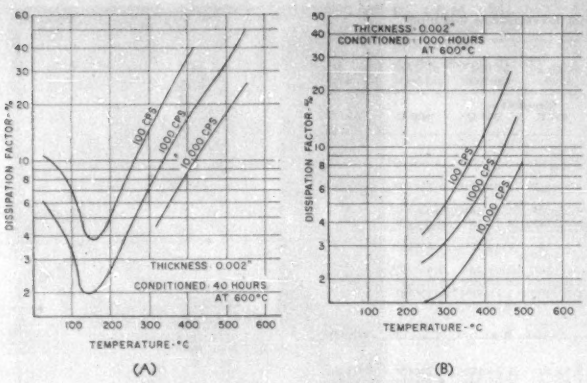


Fig. 7. Dissipation factor: silicone-bonded integrated phlogopite mica

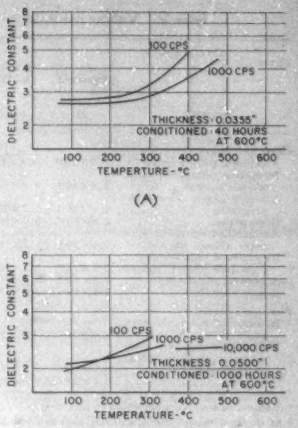


Fig. 10. Dielectric constant: aluminum-phosphate-bonded recombined muscovite mica

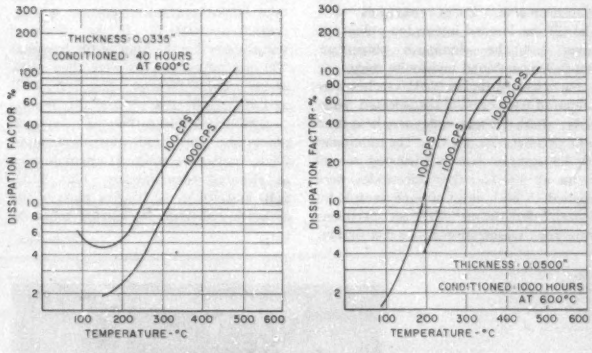


Fig. 8. Dissipation factor: aluminum-phosphate-bonded recombined muscovite mica

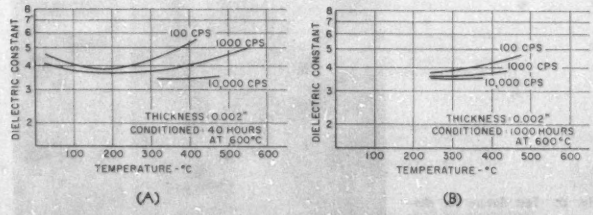


Fig. 9. Dielectric constant: silicone-bonded integrated phlogopite mica

The effects of temperature and frequency on the dissipation factor of the materials tested are shown in Figs. 7(A), 7(B) and 8(A), 8(B). The dissipation factor in general increased with temperature and decreased with frequency. The dips in Figs. 7(A) and 8(A) probably are caused by moisture in the test samples. At about 150 C the moisture is liberated from the samples and the dissipation factor then increases with increased temperature.

The effect of thermal aging can be seen by comparing Fig. 7(A) and 7(B) and Fig. 8(A) with Fig. 8(B). The dissipation factor of the silicone-bonded integrated phlogopite mica decreased during thermal aging while the dissipation factor of the aluminum-phosphate-bonded recombined muscovite mica increased.

Figs. 9(A), 9(B) and 10(A), 10(B) show the effects of temperature and frequency on the dielectric constant of the materials tested. In general, the dielectric constant increased with temperature and decreased with frequency. Again the dip in Fig. 9(A) can be attributed to moisture in the test sample. Thermal aging tends to lower the dielectric constant of both materials.

The thickness of the aluminum-phosphate-bonded recombined muscovite mica increased about 40% during heat aging (1,000 hours at 600 C). This increase in thickness could have been caused by a general delaminating within the samples. However, the sample appeared to be in

good condition physically. Delamination might also explain the erratic dielectric constant results shown in Fig. 10(B).

Conclusions

There are commercially available materials which are potential sheet and spool insulations for electronic power transformers operating in 500 C ambient temperature.

As a result of this investigation silicone-bonded integrated phlogopite mica and aluminum-phosphate-bonded recombined muscovite mica were recommended for futher evaluation as sheet insulation and spool insulation respectively in electronic transformers.

References

1. High Temperature (600°C) Capacitor, Roger L. Faust, G. W. Monk, David William Moore. *Proceedings*, National Conference on Aeronautical Electronics, Institute of Radio Engineers, New York, N. Y., May 1957, pp. 278–85.

Discussion

Arnold H. Scott (National Bureau of Standards, Washington, D. C.): In Fig. 2 of the electrode system used to determine electric strength, I do not see where the samples are. Can these be pointed out to us?

The authors used the electric strength electrodes for volume resistivity measurements, and thus obviously used no guard ring. Was the effect of conductance through the heated air determined? Mr. H. K. Brandt pointed out the difficulty resulting from conductance through the air at high temperatures in a recent report to Subcommittee V on Ceramic Products of Committee D-9, ASTM. Also, a recent visitor to the National Bureau of Standards was having difficulty with this problem. Mr. Halleck's guard ring holder could probably be used in order to eliminate this problem.

G. I. Duncan and M. M. Felger: Fig. 2 shows the electrodes used for making the electric strength and volume resistivity measurements at temperature. Insulation samples 6 inches by 6 inches were inserted between the electrodes during the tests. There are no insulation samples shown in Fig. 2.

No attempt was made to measure the conductivity of the heated air. Calculations have shown that the conductivity of air compared to that of the dielectric sample would be negligible at 600 C. Any future volume resistivity measurements will be made with the dielectric sample holder described in another paper.[1]

References

1. Sample Holder for Thin-Sheet-Insulation Dielectric Measurements Up to 600 C, M. C. Halleck. *AIEE Transactions*, vol. 77, pt. III, June 1958, pp. 343–48.

Design Considerations for Aircraft Relays

L. E. MASSIE
MEMBER AIEE

E. F. KOTNIK
MEMBER AIEE

Synopsis: The missions specified for high-performance military aircraft have continually increased the relay switching requirements. Concurrently, the environmental and operational parameters tax the maximum capability of the state of the art. In addition, these relays must have increased reliability since their purpose is to provide simultaneous multiple functions, and also controls beyond the capability of the human operator.

THE SWITCHING RELAY as presently applied in high-performance aircraft installations has been pointed to as an item of questionable reliability. It must be acknowledged that this reputation is, in part, the result of improper selection and application. The requirements have necessitated new design considerations not previously encountered. These have resulted in failures which have contributed to the further questioning of its reliability. Discussions in this presentation will be limited to specific design considerations intended to meet the latest operational and environmental conditions.

It is not reasonable to expect that all engineers and designers responsible for the selection of relays are relay or switch specialists, familiar with the many and varied applications and design requirements. Some engineers and designers have accumulated considerable reliable data by participation in, or having access to, extensive test programs involving relay installations. Others, not so fortunate, attempt to rely on vendor data and applicable "Military" specifications.

Typical relay specifications such as *MIL-R-6106* and *MIL-R-5757* can only be used as references which are intended to insure a certain minimum of quality by defining current or past performance requirements, whereas most current and future installations exceed the limits of the "*MIL*" specifications. Recognizing these limitations, much technical talent is being devoted to resolving the problems of proper relay evaluation and standardization. The results of some of this activity is becoming evident in papers such as those presented at the National Conferences on Relays at The Oklahoma Agricultural and Mechanical College under the auspices of Prof. C. F. Cameron.

Furthermore, results of the efforts of various technical groups of the Aircraft Industries Association, AIEE, Society of Automotive Engineers, and military agencies are being widely distributed for industry usage.

Description

Illustration of some of the basic considerations necessary to assure a functional and reliable switching device is presented in the following.

Once a need is established, the relay requirements must be described in sufficient detail in a document known as a procurement specification or specification control drawing, to insure adequate performance of the relay for the application intended. In addition, a test procedure must be provided that describes adequate tests to be performed which will disclose whether or not the unit will meet the performance requirements under all expected environmental conditions during its useful life. The specification should be so written that it does not describe a specific vendor's unit in such a manner that it precludes other vendors from supplying an equivalent unit from the information listed.

Paper 57-803, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, May 7–9, 1957. Manuscript submitted February 15, 1957; made available for printing March 26, 1958.

L. E. Massie and E. F. Kotnik are with Convair, Division of General Dynamics Corporation, San Diego, Calif.

A typical airframe installation requirement with the prescribed environmental conditions is

Altitude: 0 to 75,000 feet (unpressurized)
Temperature: −65 to 300 degrees Fahrenheit
Vibration: 15 to 500 cycles per second of 10 *g*
Acceleration: 10 *g* (all planes)
Shock: 50 *g* (all planes)
Acoustical noise (white noise): 50 to 8,000 cycles per second, 150 decibels
Power requirements (over temperature range): −65 to 300 degrees Fahrenheit
Nominal operating voltage: 28 volts d-c
Emergency operation: 14 volts d-c
Contacts: double-pole double-throw (form *2c*) 5 amp (ampere) 28 volts d-c and double-pole double-throw (form *2c*) 0.075 amp 28 volts d-c

Design Considerations

After requirements are established, the designer should review the following pertinent data if he expects to obtain a relay to meet these requirements.

Enclosure (Construction)

At the present time, only sealed relays can meet the prescribed environmental conditions within allowable space and weight limitations. The designer, therefore, must be concerned with the method of sealing. The term "hermetic" usually means an inorganic seal such as metal to glass or metal to ceramic. Since sealing is a matter of degree, the specification should define both the allowable leakage rate and the method of testing. In some instances, a reference to a particular method of testing will discriminate against the small manufacturer so a choice of equivalent methods should be offered. Where it is possible to examine representative samples of competitive units, the method of final enclosure should be studied, since certain solder seals now supplied are definitely inferior to others. It is not unusual for poor soldering techniques to leave small resin dikes which eventually dissolve and result in serious leaks. If solder is used, the method must not allow solder or flux to flow into the enclosure. If the mounting lugs are part of the header assembly, strength allowance must be made for tin-lead solders since they lose 50% of their strength at 100 degrees centigrade.

Terminals

In general, the sealed relays use terminal headers of the fused metal to glass type. The glass material often used is a powdered frit, preformed into pellets, and fired in an oven. It is very difficult to remove all the water gas from the frit; hence the resulting glass button has some bubbles. Where these bubbles are near, or on, the surface, the glass is rough and pitted. As received from the manufacturer, the relay header is relatively clean and can withstand the required dielectric tests. But after installation, when soldered lead wires are used, there is a tendency for the solder flux to contaminate the surface of the seal and act as the base for an accumulation of dirt. Most fluxes have appreciable conductance, and in conjunction with the accumulated dirt, lower the insulation resistance of the seal. The contaminated glass cracks easily under arcing and cannot be cleared.

This type of terminal is vulnerable to variations in the fusing operation and quite often will reveal a weakness under torsion. In some cases, the pressure of a hot soldering iron has been sufficient to release the bond and twist the terminal shaft. Besides causing leakage, this twisting can short-circuit internal contacts that are fastened directly to the terminal shaft. To eliminate this deficiency, the relay procurement specification should specify tension and torsion load limits for terminals.

Inspection of production metal to glass seals presents another problem not generally solved by relay manufacturers. Fine cracks in the glass are common. When they are less than 4 millionths of an inch in width they are not readily visible, even through a high-powered glass.[1] While these cracks may not leak initially, they can grow, through stress, into leaks.

There is an increasing trend for designers to specify relays with leads attached to the header and potted. This arrangement protects the terminals from dirt and simultaneously provides greater creepage distance for corona elimination.

Contact Ratings

1. One of the difficult problems is the selection of relays with the proper contact capacity and characteristics for the switching function. This is understandable, considering how often the designer does not have sufficient quantitative data to define his requirements.

Too often, lamp loads are rated at their nominal current and the cold filament inrush load (which is as high as 15 times the nominal load) is over looked. Contacts rated at a 50-amp resistive load have been known to melt and weld from the inrush of a 25-amp lamp load. This is due to the contact size and type of contact-make characteristic.

Inductive load ratings are even more susceptible to incorrect analysis. In one instance, a 10-amp contact relay, switching the primary of a 200-volt 400-cycle 1.8-kva (9-amp) transformer, failed in service. Oscillographic recordings revealed an instantaneous value of inrush current as high as 70 amp over a half-cycle.

Caution must be exercised in selecting contact ratings for inductive loads since variable factors (such as the L/R ratio or power factor of the load, the frequency of operation of the relay, the type of mechanical operation of the contacts, and the amount of bounce in the contacts) can result in a deficient installation. It is certainly not advisable to rely on an unspecified overload capacity of the relay contacts as a safety factor for marginal installations. If in doubt, the safest policy is to make tests under design load and design environmental conditions for the full intended life.

2. Even though all the problems of basic contact rating are satisfied, the designer is still faced with questionable electrical continuity associated with "contact contamination" and "dry circuit" switching.

Of all the problems introduced by the necessity for environment-free sealed relays, contact contamination has been the most difficult to resolve. Basically, contact contamination consists of a material deposit of varying thickness on the surface of the contact, which results in a high-resistance barrier between contact surfaces. This barrier is principally a deposit of carbon, believed to be formed by the breakdown of hydrocarbon gases in the arc between the contacts. The source of these gases is the organic materials in the sealed enclosure, as well as the impurities within the gas used for inerting. The amount of contamination is related to the operating temperature of the relay and the capacity of the organic insulating materials to contribute to carbon-forming gases. All commonly used contact materials are subject to this contamination. Materials of the platinum family have a catalyic action which aggravates the condition.[2]

Relay failure resulting from contact contamination is more often noted when contact loads are less than 0.5 amp on units of 10-amp rating. In a series of tests conducted in 1955 on type *AN-3310-1* relays of various manufacture, all relays exhibited contact resistance in excess of 0.05 ohm within the first 3,000 operations, and most contacts failed completely within 6,000 operations.[3] Since this problem would not become apparent until the relay was in service

for some time, it is mandatory to incorporate into the qualification test a procedure capable of revealing any tendency toward contact contamination. The requirement for a minimum current rating might include the following:

Resistive load: 5% of rated current or less
Inductive load: 3% of rated current
Lamp load: 3% of rated current

These tests are to be conducted at *maximum* ambient operating or storage temperature over a suitable number of cycles. Preferably the full expected life cycle should be covered, since the phenomena are accumulative.

The so-called dry circuit failure, or the difficulties of switching at such a low current value as 1 microampere at 0.1 volt, can occur without organic contamination. Where organic contamination is not a factor, the presence of metallic films such as oxides and sulfides can evoke a most perplexing assortment of troubles in relay switching circuits. These are the result of coherer action, electrostatic oscillation, electrical noise, and other phenomena.[4,5] For the present, designers can only be alerted to the possibility of these troubles. Many progressive relay manufacturers are aware of the problem and through extensive research programs are searching for a solution. It should be noted that where some contacts have been plated with special metals to reduce these films, the users' receiving inspection techniques have burned off the plating with excessive continuity checking current (usually lamp loads). It is therefore required that dry circuit relays receive special consideration to prevent damage.

In order to achieve design expectations and relay reliability, the evaluation and qualification tests must obviously be comprehensive. In addition, the tests for pickup and dropout voltage, vibration, shock, and chatter should be conducted after humidity cycles, to determine the quality of insulation.

References

1. Testing Glass Seals. H. N. Staats. *National Conference on Relays*. Oklahoma Agricultural and Mechanical College, Stillwater, Okla., 1955.
2. Switching Relay Design (book). R. L. Peek, H. N. Wagar. D. Van Nostrand Company, Inc., Princeton, N. J.
3. Report Upon the Investigation of the Contact Contamination Problem. Arthur E. Wood. *Report no. 152*, Leach Relay Division of Leach Corporation, Los Angeles, Calif., May 7, 1956.
4. Electrical Contacts (book). G. Windred. MacMillan and Company, Ltd., London, England, 1940.
5. Electrostatic Oscillation in Sensitive Contact Mechanisms. Thomas R. Stuelpnagel. *AIEE Transactions*, vol. 71, pt. II, 1952 (Jan. 1953 section), pp. 397–401.

Magnetic Circuit Properties of the Flux-Switch Inductor Alternator

S. E. RAUCH
MEMBER AIEE

Synopsis: The importance of the flux-switch-type inductor alternator as a device to provide electric power from high-shaft-speed turbine-driven sources has provided the need to extend the previous information available on the characteristics of the typical magnetic circuit used in such alternators. The experimental program presented in this paper includes data on the flux distribution in the electromagnet, yoke, stator pole tips, and rotor sections as functions of the field excitation. Standard ballistic galvanometer test procedures were employed to measure flux densities. Leakage patterns associated with the magnetic circuit are described. The relatively high internal impedance of the flux-switch alternator is shown to be the result of multiple flux paths linking the armature.

THE flux-switch alternator is a type of inductor alternator which has received attention in special air-borne applications where reliability, minimum weight and volume, and high shaft speeds are requisites.[1–3] The flux-switch alternator employs a rotor consisting of a shaft, bearings, and laminated steel punchings. As a consequence, the rotor shaft speed is usually limited by the bearing ratings. The simplicity and ruggedness of the rotor are major assets of the flux-switch alternator. The varying reluctance in selected paths of the magnetic circuit permits a reversal of flux through a portion of the iron and appropriately placed armature windings, which leads to a reduction in the volume of iron in the magnetic circuit where the flux is reversed.

As is the characteristic of all inductor-type alternators, the flux developed by armature ampere-turns in the flux-switch alternator is not restricted to the main field flux yoke, and therefore the design techniques for minimizing the internal impedance are complicated. The use of shielding, short-circuit currents, and other control devices is only partially effective in minimizing the impedance. It becomes necessary as a result to associate a relatively high impedance with the inductor-type alternator. This property is recognized as a dynamic characteristic which makes difficult the optimum use of the alternator for widely changing load conditions. The simplicity of manufacture and reliability of components for high shaft speeds, however, makes the inductor alternator a promising device for producing high-frequency single-phase power from such speeds, in particular for applications having severe environmental requirements with regard to temperature, vibration, shock, weight, and volume.

It is advantageous to examine in more detail the special configuration of magnetic circuitry which is typical of flux-switch units. In the following there are contained experimental evaluations on the flux pattern leakage factor, and inductance. The results are correlated with the dynamic characteristics of a sample alternator.

Nomenclature

ϕ_{em} = flux in electromagnet section
ϕ_p = flux in mated stator pole tip
ϕ_s = flux through saddle windings
ϕ_u = flux in unmated stator pole tip
ϕ_y = flux in yoke section
ϕ_1 = flux in unmated pole end of electromagnet section
ϕ_2 = flux in mated pole end of electromagnet section
ϕ_3 = flux in unmated pole end of yoke section
ϕ_4 = flux in mated pole end of yoke section
C = load capacitance
E_T = terminal voltage
E_g = generated voltage
i_f = field current
i_a = armature current
L_t = inductance of two windings in series
L_p = inductance of two windings in parallel
M = mutual inductance
N_f = turns in field windings

Paper 58-768, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted April 5, 1957; made available for printing March 21, 1958.

S. E. Rauch is with Sundstrand Turbo, Pacoima, Calif., and the University of California, Goleta, Calif.

Table I. Flux Distribution, Kilomaxwells, in the Magnetic Circuit of the Flux-Switch Alternator for Two Radial-Air-Gap Magnitudes and for a Rotor Position in Which a Pair of Rotor and Stator Poles Are Aligned

i_f, Amperes	Ni_f, Ampere-turns	ϕ_{EM}, Kilogauss	ϕ_{MP}, Kilogauss	ϕ_{UP}, Kilogauss	ϕ_S, Kilogauss	ϕ_{UP}, Kilogauss	ϕ_{MP}, Kilogauss	ϕ_{UP}, Kilogauss	ϕ_Y, Kilogauss
				0.0045-Inch Radial Gap					
0.050	100	4.5	4.6	4.5	1.8	1.6	1.4	5.1	2.0	3.1
0.100	200	8.8	9.2	8.7	3.5	3.1	2.7	10.1	4.0	6.2
0.150	300	12.4	13.0	12.2	4.7	4.0	3.3	13.7	5.5	8.3
0.200	400	15.1	15.6	14.9	5.5	4.6	3.5	16.0	6.7	9.2
0.250	500	16.3	17.5	16.0	5.9	4.9	3.7	17.2	7.5	9.8
0.300	600	16.4	18.8	16.1	6.1	5.0	3.9	18.0	8.0	10.0
				0.009-Inch Radial Gap					
0.030	60	1.95	2.05	1.93	0.78	0.53	0.30	1.95	0.95	1.0
0.050	100	3.2	3.35	3.12	1.25	0.85	0.58	3.15	1.35	1.8
0.075	150	4.8	5.15	4.65	1.88	1.30	0.80	4.85	2.05	2.8
0.100	200	6.5	7.0	6.45	2.87	1.77	1.10	6.6	2.85	3.75
0.150	300	9.4	10.1	9.33	3.75	2.6	1.58	9.55	4.10	5.45
0.200	400	12.6	13.4	12.3	4.88	3.35	2.00	12.5	5.50	7.05
0.250	500	14.7	15.8	14.4	5.75	3.85	2.18	14.6	6.5	8.1
0.300	600	16.4	18.0	16.1	6.4	4.23	2.4	16.3	7.5	8.75
0.350	700	17.5	18.8	17.1	6.7	4.32	2.3	17.0	8.0	8.95
0.400	800	18.3	19.7	17.9	6.8	4.4	2.2	17.5	8.1	9.30
0.500	1,000	19.8	21.1	19.3	7.0	4.3	1.93	18.6	9.3	9.25

N_s = turns in saddle windings
R = load resistance
R_{eff} = total effective reluctance
R_{gap} = air-gap reluctance
r = armature resistance

Open-Circuit Measurements

Flux

An experimental analysis has been made on a sample configuration for which the cross section is shown in Fig. 1. A stack length of 0.75 inch consisted of laminated punchings made from 5-mil Armco Tran-Cor T steel. The rotor length exceeded 0.75 inch in the test unit; therefore, the analysis of saturation effects was limited principally to stator iron sections. The laminations were punched, stacked, baked at a temperature of 400 degrees Fahrenheit, and then center-ground to final dimensions. The steel did not undergo a final annealing process in order that its characteristics might be returned to the original specifications.

A standard ballistic galvanometer measurement of magnetic flux was performed to provide an evaluation of the flux densities in the various portions of the magnetic circuit. The test circuitry employed is shown schematically in Fig. 2. The field windings were connected to a d-c supply in such a manner that the two magnetomotive forces opposed each other. Test coils which were placed at strategic positions in the magnetic circuit were successively connected to the ballistic galvanometer as shown in Fig. 2 (A). For the open-circuit measurements, the direct currents supplied to the armature windings were zero. The armature windings can be positioned as shown in Fig. 1 either as yoke- or saddle-type windings.

Table I presents a brief summary of flux magnitudes in units of kilomaxwells as a function of the exciting ampere-turns in the field windings for the rotor position in which a pair of rotor and stator poles are in line, which is the condition for maximum open-circuit flux linking the armature windings. Fig. 3 is the corresponding graphic presentation of flux magnitudes for a flux-switch alternator having a radial air gap of 4.5 mils. Linear characteristics for all portions of the iron were observed up to a 250-ampere-turn field excitation range of 300 to 500 ampere-turns. It must be noted that a saturation of the flux linking the saddle and yoke windings existed although the flux densities in the stator pole sections continued to increase with increasing field excitation. The relationship

$$\phi_Y = \phi_P - \phi_U \qquad (1)$$

is clearly shown in the data of Table I and Fig. 3.

The significance of radial air-gap magnitude can be seen from the data recorded in Table I, in which corresponding flux measurements for a 4.5- and 9.0-mil radial gap are compared. A doubling of the magnetomotive drop in the air gap produces a 10% reduction in the effective flux linking the saddle and yoke windings in the saturated regions. The flux in the electromagnet section, on the other hand, decreases only a negligible amount. The analysis in a later section shows that the iron of the electromagnet and stator mated pole sections requires large percentages of the total excitation ampere-turns, and accordingly small variations in air-gap length plays a secondary role in the storage of magnetic energy in the magnetic circuit for field excitations exceeding 400 ampere-turns.

Flux linking the saddle and yoke windings as a function of the rotor position was studied experimentally with the use of an accurate dividing table. A series of flux

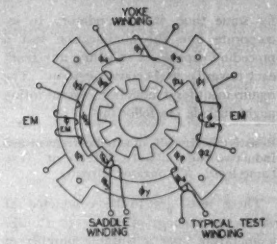


Fig. 1. Cross-section schematic of flux-switch alternator showing position of test coils, saddle, yoke, and electromagnet windings

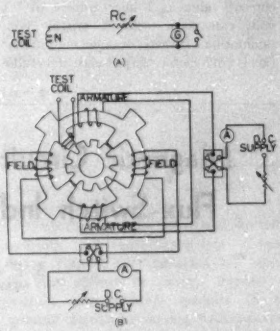


Fig. 2. Test circuitry for measuring flux using a ballistic galvanometer

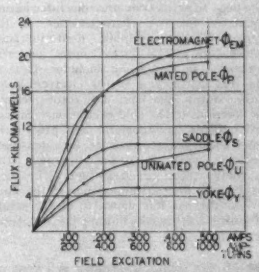


Fig. 3. Flux versus field excitation for sections of the magnetic circuit in a 0.0045-inch radial-air-gap flux-switch alternator

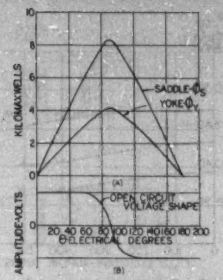


Fig. 4. (A) Flux linking saddle and yoke windings versus rotor position, θ, in electrical degrees. (B) Characteristic open-circuit voltage waveform

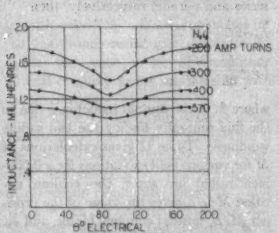


Fig. 5. Typical inductance for varying rotor position and field excitation of a 50-turn saddle winding in the 0.0045-inch flux-switch alternator

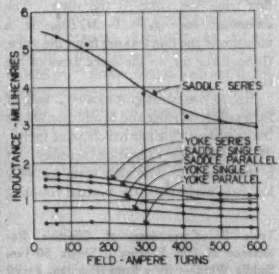
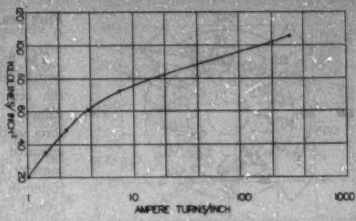


Fig. 6. Comparison of minimum inductance of 50-turn saddle and yoke windings versus field excitation in the 0.0045-inch flux-switch alternator, $\theta = 90$ degrees

measurements was obtained for armature windings for varying field currents and rotor positions. All data showed the same characteristic function of flux versus θ, electrical degrees, and for all positions of the rotor the relationship between

Fig. 7. D-c magnetization characteristic of the 5-mil Tran Cor T steel laminations used in the test flux-switch alternator



flux linking the saddle and yoke windings was

$$\phi_y = \phi_s/2 \qquad (2)$$

Fig. 4(A) illustrates a typical flux versus θ measurement. A field excitation of 320 ampere-turns and a radial air gap of 4.5 mils were used for the alternator. The electrical angle, $\theta = 90$ degrees, corresponds to a rotor-stator combination as shown in Fig. 1, that is, a pair of rotor and stator poles in line. The approximate linear relationship between flux and θ for the range 0 to 60 degrees indicates that the voltage waveform has an approximate peak to rms factor of 1.10. A characteristic open-circuit voltage shape to be expected from the flux measurements of Fig. 4(A) is shown in Fig. 4(B).

Inductance

The next section, which is concerned with open-circuit measurements, presents information on the effective internal impedance of the sample flux-switch alternator. Fig. 5 shows typical inductance values for a single-saddle winding in a flux-switch alternator with a 4.5-mil radial air gap. The use of an impedance bridge and dividing table permitted the evaluation of inductance as a function of the rotor position. The cyclical behavior of inductance with θ can be observed, and it is important to note the relatively small fluctuations in inductance in comparison with the average rms magnitude of inductance.

An interesting comparison between yoke and saddle windings can be observed from the data in Fig. 6. Fifty-turn windings, singly, in series, and in parallel, were tested to compare their effective inductance when the rotor was located in the $\theta = 90$-degree position. As a consequence of the rotor alignment with the stator, minimum inductance values were recorded. In the region of saturation the yoke windings behave as two independent windings without mutual coupling, for as can be seen from the graphs, yoke windings in series yield inductance values which are the direct sum of the inductances of the single windings. This can be expected for the paths of least reluctance, as seen by the flux linking the yoke windings, are through the two stator poles adjacent to the yoke and the rotor. Two saddle windings in parallel show an inductance which is lower than a single saddle inductance. This phenomenon reflects the existence of a basic field flux which enters at the mated pole but returns to the electromagnet section through the unmated pole surface area.

Open-Circuit Energy Distribution

An estimate of the energy storage in the various portions of the magnetic circuit requires a measurement of the d-c magnetization curve for the particular iron used in the laminated punchings. For convenience the stator stack without the rotor was excited in such a manner that the two field-winding ampere-turns were additive. With a known field ampere-turn value and measured length of iron path, the field was collapsed and the flux linking a test coil on the stator ring was measured with a ballistic galvanometer. The results of the measurements are recorded in Fig. 7. A comparatively low saturation induction of the iron occurred because the quality of steel was impaired by the processes of working the iron.

Fig. 8 summarizes the more important

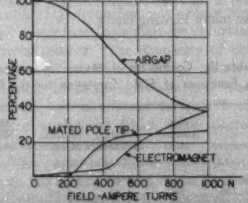


Fig. 8. Percentage distribution of energy in the magnetic circuit versus field excitation in the 0.0045-inch test alternator

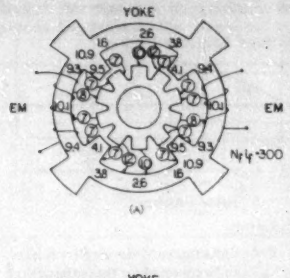


Fig. 9. Flux leakage in kilomaxwells in the 0.009-inch radial-air-gap flux-switch alternator, $\theta = 90$ degrees

A—Linear region; field excitation, 300 ampere-turns

B—Saturated region; field excitation, 600 ampere-turns

percentages of energy distribution in the magnetic circuit for the air gap, electromagnet, and mated pole tip. The energy involved in the yoke and unmated pole tip was found to be negligible. The energy in the rotor section for the test unit was insignificant also because of the extended rotor stack length. A comparison of the data in Fig. 3 and Fig. 8 correlates the start of saturation effect with the sudden percentage increase of energy in the mated pole tip. With increasing field excitation it is shown that the saturation of the electromagnet section becomes the major limiting factor.

Flux Leakage

The flux leakage occurring in a flux-switch alternator significantly affects the generated voltage and inductance values. Leakage occurs principally in three sections: electromagnet, yoke, and unmated stator poles. For example, consider the two field excitations illustrated in Figs. 9(A) and 9(B) which are typical for the 0.009-inch radial-air-gap alternator operating in the linear and saturated regions. The flux leakage magnitudes are shown encircled. In either case, one can note a very significant leakage from the electromagnet and yoke sections to the rotor. In Fig. 9(A), there exists an approximate 0.8-kilogauss leakage in the electromagnet which does not link any turns of a saddle winding. The unmated pole shunts 5.5 kilogauss from the main field flux; this shunted flux must be subtracted from the flux through the mated pole in order to obtain the effective maximum flux linking a saddle winding. As a result, a total 6.3-kilogauss leakage occurs from the total 10.1 kilogauss developed in the electromagnet. In the yoke section 2.2 kilogauss leak to the rotor section and can link at most only one saddle winding; consequently an effective loss of at least 1.1 kilogauss occurs in the yoke section, thereby providing a total equivalent of 7.4 kilogauss. The complete reversal of flux through the saddle winding compensates the effective linkage of $10.1-7.4=2.7$ kilogauss to produce a net 5.4-kilogauss flux change per saddle winding. The ratio of effective flux linkage to generated maximum flux in the electromagnet section is $5.4/10.1=0.53$.

Similarly, for Fig. 9(B), which is typical of the saturated operation, the ratio of effective flux to maximum flux in the circuit is $8.75/18=0.49$.

Further evidence on flux leakage is available from an examination of the inductance measurements on saddle windings singly, in series, and in parallel. It is known that if L is the inductance of a single winding, and if L_s and L_p are the inductance of two identical windings in series and parallel respectively, then

$$L_s = 2L + 2M \tag{3}$$

$$L_p = \frac{L^2 - M^2}{2L - 2M} \tag{4}$$

where M is the mutual inductance due to the flux mutually linking the two saddle windings. Table II gives calculations of M for various field excitations for a 0.009-inch radial air gap. The column entitled $M\phi_s/L$ gives the value of the product M/L, mutual coupling factor, and ϕ_s, flux linking a saddle winding. Similarly, column $(1-M/L)\phi_s$ estimates the magnitude of the flux which links one saddle winding but which does not mutually link both saddle windings. The final column, $\phi_u - \phi_p$, records the leakage of flux from the yoke to the rotor. The difference $(\phi_u - \phi_p) - (1-M/L)\phi_s$ is a measure of the flux leaking from the yoke, through the air, to the stack faces of the rotor.

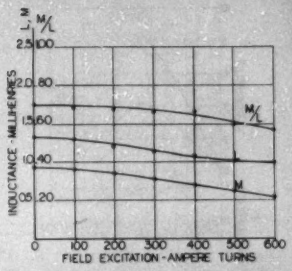


Fig. 10. Comparison of inductance, mutual inductance, and mutual coupling factor of saddle windings in the 0.009-inch flux-switch alternator

Fig. 10 compares graphically L and M for a saddle winding and in addition shows the decrease in the mutual coupling factor, M/L, as the field excitation increases. It is important to note that even for very low field excitations the mutual coupling

Table II. Coefficient of Mutual Inductance and Mutual Flux Linking Saddle Windings as Functions of Field Excitation for the 0.009-Inch Radial-Air-Gap Flux-Switch Alternator

i_f, Amperes	$N_f i_f$, Ampere-turns	L (Single), Millihenrys	L_s (Series), Millihenrys	M, Millihenrys	M/L, Percentage	ϕ_s, Kilogauss	$M\phi_s/L$, Kilogauss	$(1-M/L)\phi_s$, Kilogauss	$\phi_u - \phi_p$, Kilogauss
0.050	100	1.30	4.4	0.90	0.69	1.8	1.2	0.6	0.67
0.100	200	1.23	4.15	0.84	0.68	3.6	2.4	1.2	1.47
0.150	300	1.15	3.84	0.77	0.67	5.4	3.6	1.8	2.17
0.200	400	1.07	3.55	0.70	0.65	7.1	4.5	2.6	2.88
0.250	500	1.03	3.30	0.62	0.60	8.1	5.0	3.1	3.57
0.300	600	0.99	3.10	0.56	0.57	8.8	5.0	3.8	4.0
0.400	800	0.95	2.95	0.52	0.55	9.2	5.1	4.1	4.6

Table III. Comparison of Effective Reluctance of Flux Linking a Single 50-Turn Saddle Winding for a Rotor Position, $\theta = 90$ Degrees, for a 0.0045-Inch and 0.009-Inch Radial Air Gap

		0.0045 Inch		0.009 Inch	
i_f, Amperes	$N_f i_f$, Ampere-turns	L_s, Millihenrys[1]	R_{eff}	L_s, Millihenrys	R_{eff}
0.050	100	1.54	0.021	1.30	0.024
0.100	200	1.42	0.022	1.23	0.026
0.150	300	1.25	0.025	1.15	0.028
0.200	400	1.10	0.029	1.08	0.031
0.250	500	1.00	0.032	0.97	0.033
0.300	600	0.97	0.033	0.95	0.034
0.400	800	0.93	0.034	0.91	0.035

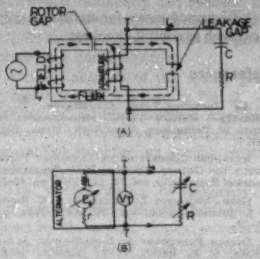


Fig. 11. (A) Equivalent magnetic circuit of the flux-switch alternator. (B) Equivalent electric circuit of flux-switch alternator including variable generated voltage

factor is only 70%. Similar estimates for the 0.0045-inch radial-gap alternator indicate the same order of magnitude mutual coupling factor relationship with field excitation.

The large values of flux leakage affect the accurate prediction of inductance if one uses the air-gap dimensions to determine the reluctance. The possibilities of erroneous results are shown in Table III, in which the effective reluctance, R_{eff}, has been calculated and recorded for a single saddle winding by the use of the relationship

$$R_{eff} = \frac{1.26 \times N_s^2 \times 10^{-8}}{L} \qquad (5)$$

where R_{eff} has the basic units of centimeter $(cm)^{-1}$. The corresponding air-gap reluctances, assuming an air-gap area of 1.03 cm^2, are 0.0229 cm^{-1} and 0.0458 cm^{-1} for the 0.0045-inch and 0.009-inch radial air gap, respectively.

The existence of parallel flux paths which link the saddle winding reduces the total reluctance in accordance with the usual reciprocal relationship as used for parallel resistances. Consequently, it is not surprising to note that the effective reluctances for low field excitations prove to be lower than the air-gap reluctances. Increasing field excitations produce increasing series reluctance in all parallel paths which contain iron because of the rapid decrease in permeability of the iron with increasing flux densities. For the 0.0045-inch radial-gap case, the effective reluctance increases rapidly and exceeds the air-gap reluctance. The 0.009-inch gap data show only slight increases in the effective reluctance over the previous example although the basic air-gap reluctance has been doubled.

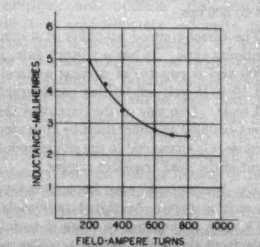


Fig. 12. Rms inductance of 50-turn saddle windings in series evaluated from dynamic series capacitance resonance tests

Generated Voltage Versus Armature Reaction

In Fig. 11 there are illustrated simplified circuit schematics for the purpose of discussing the concept and for evaluating the generated voltage under load conditions. The a-c source shown in Fig. 11(A) represents the switching of the field flux through the armature windings as a result of the shaft rotation. The field flux, therefore, is considered as a cyclical flux linking the armature windings and passing through the two rotor air gaps. A portion of the main field flux links instead the leakage circuit, shown as a parallel iron path with a very large air gap. If the armature current is so phased that the ampere-turns due to it oppose the field flux, then the effective reluctance for flux linking the saddle winding is increased. The field flux then is reduced through the saddle-winding, and a larger percentage of the flux passes through the parallel leakage paths.

The schematic in Fig. 11(B) represents an equivalent circuit for the flux-switch alternator in which the generated voltage, E_g, is varying and is a function of the phasing and magnitude of the armature reaction. L is the equivalent inductance of the armature, r is the d-c resistance of the armature windings, C and R are, respectively, the capacitance and resistance in the load circuit, and E_T is the terminal voltage of the alternator. Accordingly,

$$E_g = i_a Z$$

where

$$Z = \sqrt{(R+r)^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \qquad (6)$$

and

$$E_T = i_a \sqrt{\left(R^2 + \frac{1}{C}\right)^2} \qquad (7)$$

The effective generated voltage versus armature reaction has been computed on the basis of equations 6 and 7. Supplementary data on the rms inductance of saddle windings were obtained statically by aligning a pair of stator and rotor poles, exciting the field windings, and then resonating the armature windings with series capacitance and a-c power source which could load the armature windings. The resultant computed rms inductance values of 50-turn saddle windings in series is given in Fig. 12. The corresponding values of inductance in Fig. 6 obtained with a low-power-source impedance bridge differ by less than 10%. Fig. 13 gives a summary of a series of test runs to determine generated voltage as a function of armature reaction and field current excitations. The generated voltages recorded in Fig. 13 correspond to unity power factor. Varying the phase angle from positive to negative indicated that the generated voltage could differ by as much as minus to plus 10% from the results in Fig. 13 for a given magnitude of armature reaction and field excitation.

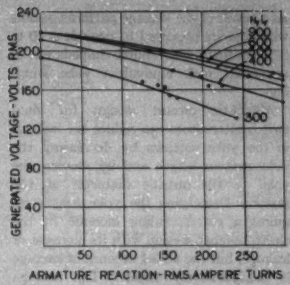


Fig. 13. Generated voltage versus armature reaction for the 0.0045-inch radial-air-gap flux-switch alternator

Conclusions

The flux-switch inductor alternator has been proved by experience to have an important role in the application of generating high-frequency single-phase electric power from high-shaft-speed turbine-driven sources. The simplicity of electrical and mechanical design makes it possible to produce reliable and maintenance-free alternators for use in airborne equipment which must operate under severe specifications.

This paper has shown that relatively high flux leakage occurs in the magnetic circuit of the flux-switch alternator. For operation near the saturated condition, the ratio of effective flux linking the armature windings to the maximum generated flux in the electromagnet is as low as 50%.

As a result of the leakage patterns, saddle windings are observed to provide a 20 to 30% reduction in source impedance as compared with equivalent yoke windings.

Magnetic circuit design for flux-switch alternators can reduce the leakage in the yoke sections by decreasing the yoke section length and by moving the yoke to the outside diameter of the stator. Decreasing the yoke length requires a corresponding increase in the electromagnet section with its increase of leakage, but the added amount of leakage is comparatively small. The leakage from the pole tips is important in determining the effective flux linking the armature windings, for the greater the leakage is in these sections, the more effective the flux reversal is through the armature windings reduced. It is the latter leakage which determines the maximum effective flux linking the armature windings.

As a result of the inherent high internal impedance due to flux leakage paths, the flux-switch alternator should be accompanied by allied regulation circuitry if large load variations are to be permitted. The flux-switch alternator is best suited for applications requiring single-phase high-frequency power for auxiliary equipment such as control circuitry.

References

1. An Inductor-Type Alternator, F. E. Steinberg. *Master's Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., Sept. 1953.

2. Advanced Concept of an Accessory Power Supply for Guided Missiles, R. L. Zimmerman. *Electrical Engineering*, vol. 75, Jan. 1956, pp. 48–51.

3. High-Frequency Alternators, S. E. Rauch, L. J. Johnson. *Ibid.*, vol. 73, Aug. 1954, pp. 735–37.

4. Design Principles of Flux-Switch Alternators, S. E. Rauch, L. J. Johnson. *AIEE Transactions*, vol. 74, pt. III, Dec. 1955, pp. 1261–68.

Frequency Transient Measurements in Power Systems

G. N. SWANSTROM
ASSOCIATE MEMBER AIEE

CONVENTIONAL instruments presently used for frequency transient measurements in power systems usually consist of a frequency discriminator, a direct-coupled amplifier, and an ink-writing oscillograph. Most frequency discriminators change calibration with changes of signal voltage and harmonic content. They exhibit drift as component parts age. Direct-coupled amplifiers categorically drift. Ink-writing oscillographs do not produce highly accurate records because of limitations in repeatability and readability. More accuracy may be obtained in measuring the magnitudes of frequency transients by using a recently developed system called the "AFTRAMAG" (*A*ccurate *F*requency *Tra*nsient *Mag*nitude) measuring system.

AFTRAMAG employs the same type of frequency discriminator, direct-coupled amplifier, and ink-writing oscillograph described in the foregoing. However, accuracy of the data does not depend on the calibration of these units. The readability of AFTRAMAG is considerably improved by digital readout.

Paper 58-788, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted February 19, 1958; made available for printing April 7, 1958.

G. N. Swanstrom is with the Sundstrand Aviation Division of the Sundstrand Machine Tool Company, Rockford, Ill.

The author wishes to express his thanks for suggestions made by S. S. Baits and E. W. Erikson, and to acknowledge the development and design work performed by R. A. Allen and J. G. Bourit.

Operation of System

The system consists of a frequency discriminator, a direct-coupled amplifier, a peak transient trigger, an electronic frequency-period counter, and either an ink-writing or photographic oscillograph. Fig. 1 shows the block diagram and typical voltage waveforms for the system. The frequency discriminator is connected to a 400-cps (cycles per second) bus where the frequency transient to be measured occurs. A voltage analog of frequency produced at the output of the frequency discriminator is amplified by the direct-coupled amplifier. The resulting waveforms may be seen at the right of Fig. 1. After amplification, the voltage analog of frequency is fed to the peak transient trigger and to the oscillograph. When the peak of the transient occurs, the peak transient trigger senses zero slope and sends a pulse to the electronic frequency-period counter. The counter adjustments are set for a period measurement and connections are made so that a signal is received from the 400-cps bus. At the instant the counter receives the pulse it resets and then measures the period of the next voltage wave on the 400-cps bus. When the counter starts counting, it displays a gate pulse on the oscillograph. By oscillographic comparison of the time of the gate pulse and the time when the peak of the transient occurred, timing of the counter trigger may be readily checked. The counter holds the period measurement until another reset pulse is presented or until it is manually reset. The period measurement may be converted to frequency by using the reciprocal relationship.

Description of Components

With the exception of the peak transient trigger the components making up AFTRAMAG are commercially available.

The accuracy of AFTRAMAG does not depend on the calibration of the frequency discriminator. Therefore, the discriminator need not have exceptional linearity or drift characteristics. Its sensitivity to 400-cps bus voltage and harmonic content changes is of little importance. The frequency discriminator used has a standard double-tuned circuit. It is tuned for zero volts output with 400-cps input frequency. The polarity of the output depends upon whether the input frequency is above or below 400 cps. Saturation does not occur until the input frequency exceeds the range of 300 cps to 500 cps.

The direct-coupled amplifier is a conventional type. It has a voltage gain of 1,000 and is capable of driving a standard 15–0–15-volt ink-writing oscillograph full scale. One side of the output of the direct-coupled amplifier must be grounded because it is connected to the single-ended input of the peak transient trigger.

The oscillograph may be a standard ink-writing type. As stated previously, ±15 volts is required to drive it to full scale from the center position. The oscillograph should have a sine wave frequency response flat to 30 cps. During very fast frequency transients, when

oscillograph readability and response are critical, it is advantageous to use a photographic oscillograph which has fast paper speeds and galvanometers.

A block diagram of the special peak transient trigger is shown in Fig. 2. It consists of a direct-coupled amplifier-inverter, a differentiator, a 2-stage capacitor-coupled amplifier, an arming circuit, and a trigger thyratron.

The voltage analog of frequency is first inverted and then differentiated as shown by the typical voltage waveforms on the right of Fig. 2. The differentiated signal is amplified and presented to the trigger thyratron. The trigger thyratron has two control grids. One is used as a signal grid and the other is used as an arming grid. The amplified differentiated signal is fed to the signal grid. As the voltage analog of frequency approaches peak value, the signal grid voltage approaches zero volts rapidly. Prior to the transient, the arming circuit holds the arming grid sufficiently negative to hold the trigger thyratron cut off.

Arming of the trigger thyratron, as the frequency transient approaches its peak, is necessary so that it will not fire at the peaks of smaller insignificant frequency transients. The smaller frequency transients occur because of normal bus frequency modulation and synchronous generator field shifts during load changes. The first peak and notch on the voltage analog of frequency waveforms of Figs. 1 and 2 are examples of the frequency transients caused by synchronous generator field shifts.

After the transient voltage analog of frequency raises the arming grid from -6 volts to -2 volts, the signal grid controls the firing of the thyratron. In order for the signal grid to obtain more accurate control, the arming grid voltage is held constant at -2 volts during the extreme part of the transient. A Zener reference diode is used for this purpose. The trigger thyratron fires when the following conditions are fulfilled: the arming grid is at approximately -2 volts, the signal grid is at approximately -1 volt, and the plate voltage is at $+230$ volts. By inspecting the typical voltage waveforms of Fig. 2, it can be seen that the reset trigger pulse is generated near the peak of the frequency transient. To insure proper arming circuit operation it is necessary for the voltage analog of frequency, which is fed to the peak transient trigger, to be between $+10$ and $+20$ volts in magnitude. A simple adjustment of the direct-coupled amplifier gain control is required to set this magnitude.

The circuit diagram for the peak transient trigger is shown in Fig. 3. All direct voltages used in the circuit are supplied by the electronic counter. Some filtering has been added in the trigger circuit for the $+230$-volt source. The capacitors in the filter store energy to be used when the trigger thyratron fires, thus preventing a decrease of voltage in the $+230$-volt supply which might influence the operation of the counter. On the input to the trigger circuit is a zero center microammeter and multiplier resistor. The microammeter indication is used as a monitor for the system balance, transient magnitude, and transient direction. C_8 and R_7 make up the differentiator. Coupling circuits in the 2-stage capacitor-coupled amplifier have large RC time constants. In order to discharge the coupling capacitors completely, a switch SW_1 having two sets of normally closed spring return contacts is employed. The switch also has one set of normally open contacts, which when closed, connects $+230$ volts to the plate of the trigger thyratron. The operator must hold the switch down when the measurement is being made. A neon lamp is used in the plate circuit of the trigger thyratron to indicate its firing. C_9 and R_{14} differentiate the step of voltage which occurs across R_{15} when the trigger thyratron fires. This provides a pulse for the counter reset circuit. C_{10} blocks the direct voltage from the counter reset circuit.

The electronic frequency-period counter

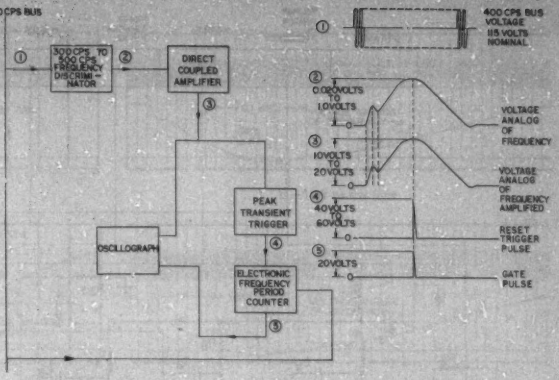


Fig. 1. Block diagram of Accurate Frequency Transient Magnitude measuring system

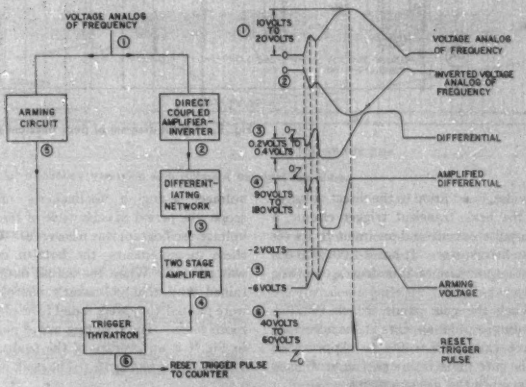


Fig. 2. Block diagram of peak transient trigger

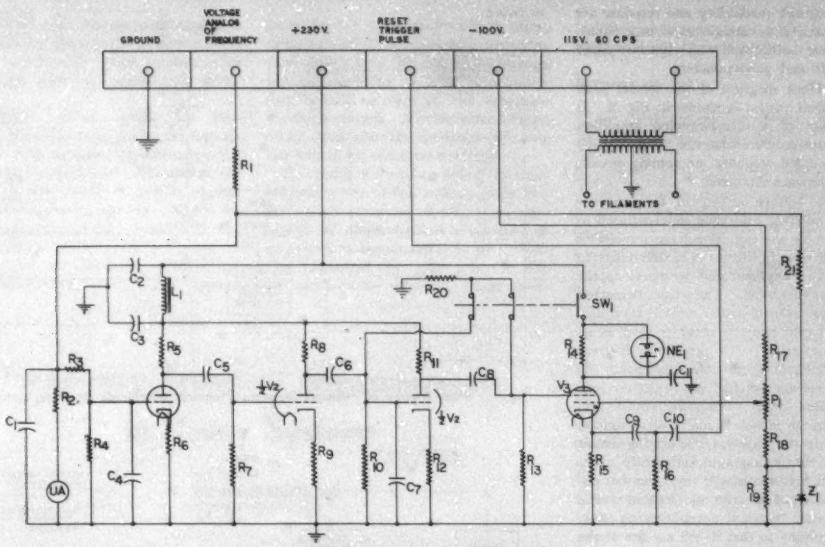


Fig. 3. Circuit diagram of peak transient trigger

provides, in addition to the direct voltages for the peak transient trigger circuit, a gate pulse output and an input to its reset multivibrator. It has a 1,000,000-cps crystal time base and six decade counting units.

When the gate opens, which initiates counting, a pulse appears at the accessory socket gate pulse contact. Appearance of the gate pulse is an exact indication of the start of the measurement.

Response

The display of the gate pulse on the oscillograph is a true indication of when counting began. By oscillographic comparison of the timing of the gate pulse with respect to the peak of the voltage analog of frequency, it is possible to check how near the peak the measurement was made. It is important that there be no appreciable phase shift or lag between the indicated peak and the actual peak of the transient.

Fig. 4 shows a frequency transient from a 400-cps rotary synchronous generator, which has been recorded on a photographic oscillograph. The voltage analog of frequency was recorded from the output of the frequency discriminator. The top oscillogram is the 400-cps bus voltage. Using a toolmaker's microscope, the period of each cycle of the bus voltage oscillogram was measured. With these measurements, the bottom curve was plotted. When the actual data obtained with the toolmaker's microscope were plotted, they were found to be of sawtooth form. The resulting curve plotted in Fig. 4 is an average of the peaks and valleys of the sawtooth. The peak of the resulting curve is a reliable indication of the true peak of the transient; it does not necessarily indicate absolute period values. The shape of the voltage analog of frequency oscillogram is not a mirror image of the bottom curve since the ordinate scales are different. The ordinate scale for the period plot was chosen so that the peak of the curve could be easily seen. The non-sine-wave appearance of the

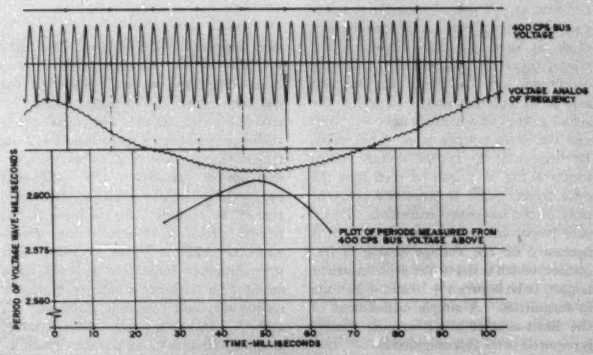


Fig. 4. Frequency discriminator response curves

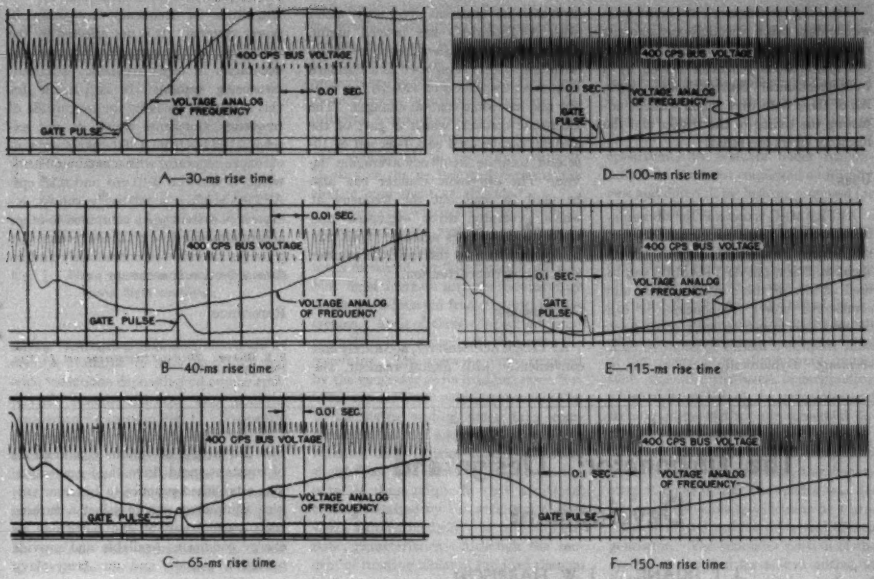


Fig. 5. Measurement of frequency transients with different transient rise times

recorded 400-cps bus voltage is due to darkening of the oscillogram with ink so that it would reproduce more satisfactorily. The peak due to synchronous generator field shift is off the left end of the figure. The rise time of the transient is about 50 ms (milliseconds). Transient rise time as used in this paper is the time from the nominal end of generator field shift effects to the approximate peak of the transient. It can be seen that the actual peak of the transient and the indicated peak are within 0.5 ms of each other. The error produced by this phase shift is so small that it is insignificant.

Since the direct-coupled amplifier has no reactive parts it can be considered to have zero phase shift. Standard available ink-writing oscillograph galvanometers have a rise time, to a step of voltage, in the order of 10 to 15 ms. Because of the inertias involved in most rotary synchronous generating systems, the fastest normal transient change from minimum to maximum is in the order of 40 to 50 ms. Suitable ink-writing oscillograph galvanometers are considered to follow this change accurately. If faster transients are encountered, a photographic oscillograph may be used to check the time of the trigger.

Accuracy

The accuracy of AFTRAMAG depends on:

1. The accuracy of the electronic frequency-period counter.
2. The ability of the system to make the measurement at the peak of the transient.

The accuracy of the counter is plus or minus one count in the last decade plus the crystal time base inaccuracy. The long-term crystal-time-base inaccuracy is three parts per million per week. Plus or minus one count, when reading the period for 500 cps, is one part in 2,000. Plus or minus one count for 300 cps is one part in 3,333. Converting these inaccuracies to frequency errors, it is found that for a transient which deviates to 500 cps, the error is 0.25 cps. For a transient which deviates to 300 cps, the error is 0.10 cps. Counter errors for other frequencies between 300 cps and 500 cps are between these limits.

Figs. 5(A)–(F) demonstrate the ability of the system to make measurements at the peaks of transients with different transient rise times. These photographic recordings were made by applying different amounts of load to a 40-kva 400-cps rotary synchronous generator. Generator drive parameters were varied purposely to cause slow and fast transients. For transients with rise times between 30 and 150 ms, the timing error results in a peak frequency transient error of less than one per cent of the actual peak frequency transient magnitude.

In using AFTRAMAG to measure frequency transients during bus fault conditions or where the transient magnitudes are larger and slower, for maximum accuracy, it is necessary to adjust the trigger timing slightly. The disadvantage of making such an adjustment may be that the fault condition must be repeated and this causes unnecessary hardship on the equipment being tested.

Table I. Frequency Transient Magnitude Data

Load Application, Per Cent of Full Load	AFTRAMAG Indicated Peak Value of Transient, Cps	Maximum Error, Cps
25	396.8	0.17
50	392.9	0.20
100	385.1	0.30
150	378.2	0.37
200	371.1	0.43

Table I gives frequency transient magnitude data taken with different amounts of step load applications on a 40-kva 400-cps synchronous generator driven by a hydromechanical constant-speed drive. All of these data were taken without adjusting the timing of the trigger and the resulting errors are listed.

Uses

AFTRAMAG is useful in the laboratory as an accurate power-frequency transient magnitude measuring device. With slight modifications AFTRAMAG can be used at any of the normal aircraft and commercial power frequencies. It can be used in conjunction with other transient frequency meters to improve their accuracy. Dynamically calibrating a conventional transient frequency meter with AFTRAMAG has proved useful. Calibrating the transient frequency meter with this system being used as a standard eliminates the effects of 400-cps bus voltage and harmonic content changes. The electronic counter, which is part of the system, can be used as a 1-second or 10-second accurate frequency-averaging device. The electronic counter can also be used to read out the frequeny at which generator drive overspeed and underspeed switches actuate. This can be accomplished by resetting the counter with the switch actuation.

Conclusions

AFTRAMAG measures, accurately and conveniently with digital readout, the magnitude of frequency-transients in power systems. It is not affected by the variables which normally produce inaccuracies in other frequency transient measuring systems. It is possible for this system to measure the magnitude of frequency transients which do not exceed ±100 cps from a 400-cps rotary synchronous generator with a maximum error which is between 0.10 cps and 0.25 cps. AFTRAMAG is a relatively simple, inexpensive system when compared to other accurate analog-to-digital converters. It provides information immediately and no data reduction is necessary.

Reference

1. 400-Cycle Transient Frequency Meter, F. E. Dickey. *Electrical Engineering*, vol. 76, Dec. 1957, pp. 1076–78.

The Turbonator: Design and Development

J. T. DUANE
ASSOCIATE MEMBER AIEE

J. W. HARRISON
NONMEMBER AIEE

THE BASIC configuration of the electric generating systems used on aircraft has been changing rapidly in recent years. This change has carried from the wide-speed-range 30-volt d-c systems used on most World War II planes to the constant-frequency 400-cps (cycle per second) a-c systems used on newer aircraft. The advent of constant-frequency requirements on airplanes with variable-speed engines necessitated the development of techniques for supplying constant-speed shaft power to the generators. In some cases this is done with shaft power supplied through variable gear ratio devices and in other cases by fixed-speed air turbines operating on high-pressure air bled from the jet engine compressor. Regardless of the approach used, the electric system designer faces two major problems; overall system weight must be minimized, and the system components must be compatible with the severe environmental conditions associated with high-performance aircraft. Both of these problems promise to become more severe in future aircraft.[1]

It is in an effect to meet these needs of the future that the Turbonator has been conceived and developed. The name "Turbonator" comes from the unique combination of turbine and generator, integrated to give a constant-frequency generating system with capabilities exceeding the sum of the capabilities of the individual components. Active design of the Turbonator was begun in mid-1956. Feasibility studies preceded the actual development of components and techniques necessary to the program. Because of the success of these preliminary efforts, the first complete Turbonator was brought to test at the beginning of 1957. It has been undergoing extensive evaluation tests under all types of operating conditions since that time. The discussion presented here is intended to describe the basic concepts behind the Turbonator and some technical problems encountered and overcome during its development.

Paper 58-791, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 23–27, 1958. Manuscript submitted March 13, 1958; made available for printing April 7, 1958.

J. T. Duane and J. W. Harrison are with the General Electric Company, at Erie, Pa., and West Lynn, Mass., respectively.

The Turbonator Concept

The initial design studies on the Turbonator were based on an analysis of the installed weight of various combinations of constant-speed drives and generators. The established practice in bleed-air turbine equipment was to select a turbine shaft speed based on the particular air energy conditions available and provide reduction gearing and an Army-Navy mounting pad for mounting a 6,000-rpm 400-cps generator.

Design studies made on both the geared turbine system and a direct-drive turbine system established that there was a substantial weight advantage to the direct-driven system without reduction gearing. Further design study clearly demonstrated the advantage of eliminating the Army-Navy drive pad and integrating the generator and drive into a single package using a common shaft for both turbine and generator with only two main bearings. This arrangement gave a system weight lower than either the hydraulic or air turbine constant-speed drive systems or the integrated hydraulic constant-speed drives, and has been given the name Turbonator. Comparative equipment weights are tabulated in Table I.

Conventional air-turbine-generator drives are ducted from the main jet engine and operate on air bled from the

Table I. Comparative Machine Weights

System	Estimated Weight
Bleed air turbine and 6,000-rpm alternator	155
Hydraulic drive and 6,000-rpm alternator	137
Integrated hydraulic drive	135
Turbonator	95

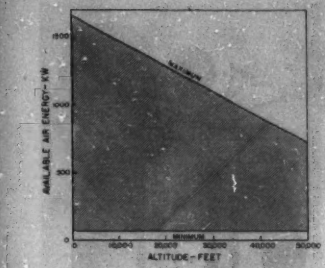


Fig. 1. Bleed air temperature and pressure versus flight condition

outlet of the engine's compressor. The energy available in this bleed air has wide variations depending on engine rpm, aircraft speed, altitude, and ambient temperature as seen in Fig. 1. To operate the turbine at peak efficiency over the energy range would require both a variable geometry and a variable speed turbine. Since the shaft output speed must be constant, the turbine designer is faced with a design compromise. In general, this compromise is made to favor the lowest energy conditions encountered. This is done to achieve highest efficiency at these points and results in the minimum duct size and minimum turbine areas. With these conditions and the required loads established it is then possible to establish the optimum turbine diameter and shaft speed. An analysis of current and proposed jet engine bleed conditions as well as electrical loads from 10 to 180 kva established that 24,000 rpm would provide a desirable standard shaft speed from turbine design considerations. The concept of an integrated package of air turbine and generator leads naturally to the consideration of cooling the generator and bearings with the air that has passed through the turbine. When a turbine is operating and delivering power, the temperature drop through it is proportional to the available energy in the air and the turbine efficiency, while the air flow is proportional to the load and inversely proportional to the efficiency. It is these characteristics which make turbines the principal source of cooling air on aircraft air-conditioning systems and could provide cooling at all conditions without blowers or the necessity of derating the electric equipment at altitude.[2]

Cooling characteristics for an integrated turbine generator system are shown in Fig. 2. As electrical load on the generator is increased, air flow through the turbine must increase and the temperature of the air exhausted from the turbine decreases. Both of these effects lead to increased cooling capacity available for the generator. This extra cooling is required by the generator as its load increases, but in the typical Turbonator, as shown in Fig. 2, the extra cooling more than compensates for the additional losses, and winding temperatures actually decrease slightly as load increases. The Turbonator is thus unique in that additional cooling required by the generator is provided automatically by the turbine's airflow characteristics. Although the concept of running cooler at full load than no load was most unconventional, the winding temperature requirements for this system appeared to be reasonable and it was decided that the Turbonator package would be self-cooled. Fig. 3 represents the complete system in block form.

Evaluation of Prototype

The basic concept of the Turbonator as described thus far looked good, but, as with any other new and radically different approach, experimental proof of feasibility was needed. In order to establish this proof, and to gain increased understanding of the detailed engineering principles involved, a complete Turbonator package was built and tested. Doing this made it possible to check by test the predicted characteristics of the 24,000-rpm generator when operated in conjunction with a direct-connected turbine. It permitted the designers to learn from experience the problems encountered in integrating the turbine and generator into one package. The construction of this development sample was an excellent opportunity to evaluate under realistically controlled test conditions a radically new and different air bearing proposed for Turbonator use.

Another important objective of the development effort was the experimental evaluation of turbine exhaust cooling. In all of these areas there was much to be learned. Construction and test of the development design provided useful knowledge of the performance characteristics of the Turbonator. Even more important was the increased understanding which was gained of the principles and techniques needed for the successful design of future models.

To provide a direct basis of comparison with conventional machines, the demonstrator model of the Turbonator was designed to the same specification as a production geared turbine drive and 60-kva generator. The generator portion of the system is designed for 40-kva output at 24,000 rpm under conventional class *C* cooling conditions.[3] It uses a static excitation system, is 22 inches long, 10¹/₂ inches in diameter, and weighs 96 pounds exclusive of controls. An external view of the unit with test instrumentation is shown in Fig. 4.

Fig. 5 pictures the physical location of the various system components. Multiple ducting at the top of the illustration is provided to accommodate the partial arc type of air supply described in the section on turbine design. The turbine wheel is immediately below the air inlet ducts and a 4-inch hydrostatic air bearing is next to the wheel. The 2-pole round-rotor generator is next on the shaft, and below it is a small 1,600-cps permanent-magnet generator. Below the perma-

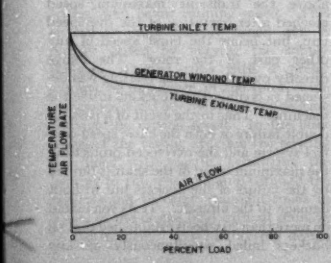


Fig. 2 (left). Cooling characteristics versus load

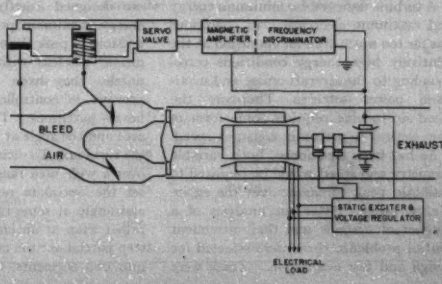


Fig. 3 (right). System block diagram

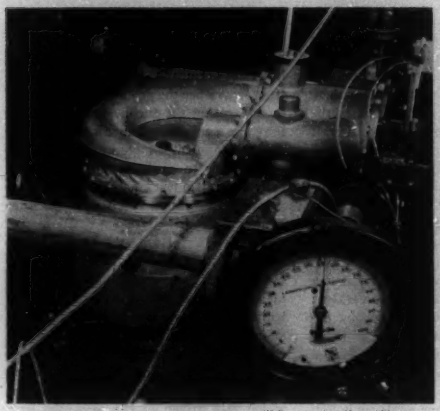


Fig. 4 (left). External view at test

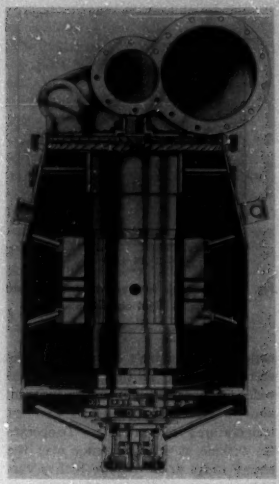


Fig. 5 (right). Section view of system

nent-magnet unit are the slip ring assembly and an oil-lubricated ball-bearing capsule. Design of the rotating parts was such that it was possible to complete the rotor assembly before inserting it into the stationary housing, thus permitting close tolerances on balance and allowing accurate measurements of balance during the course of the test program.

The isochronous speed control system used on this system receives its basic speed signal from a permanent-magnet generator located on the main shaft and produces d-c signal proportional to speed error as described in detail in a separate paper.[4] This signal is fed to a servo valve which operates two actuators, one for each control valve. The proper sequencing of the control valves is obtained by a spring bias in the power valve which holds it closed until the cruise valve is fully opened. The system has a frequency modulation error of less than $\pm 1/2\%$.

Turbine Design

A turbine designed for minimum energy and maximum electrical load conditions has far too much area for operation at the relatively high energy conditions corresponding to the aircraft cruise and maximum power settings. Therefore the bleed-air turbine requires some form of flow control to maintain constant speed.

As has been mentioned, both variable geometry and speed would be required to maintain peak efficiency over the entire energy range. After an analysis of a number of systems and their attendent control problems, three were selected for design and test evaluation. These were throttling, which in effect reduces the available energy to a point where it matches the applied load; variable-angle nozzle, which reduces the flow area to match the applied load; and a 2-step partial arc which acts as a combination of both of the other two. The experience gained in the design and manufacture of the three inlet systems established that the throttling system was the lightest and the least expensive. The 2-step partial arc system was a close second while the variable-area nozzle system was the heaviest and most costly to produce. The testing program established that the throttling and partial arc systems gave equal performance at the lowest energy conditions and superior performance to the variable-angle system. At the design cruise point the partial arc system was superior to both the throttling and variable-angle systems while at points between cruise and idle engine conditions, the variable-angle system was superior.

Perhaps some explanation of these results is in order. With the throttling system the turbine buckets and nozzles are designed exactly to the minimum energy maximum power point and it is possible to peak the efficiency at this condition, whereas with the variable-angle nozzle the shape of the nozzle air passage is controlled by rotating the nozzle partitions. This varies both the area and the angle at which the air strikes the turbine. To maintain good efficiency over a wide area range it is necessary to set the optimum nozzle passage shape and angle at some intermediate position rather than at an end point. In the 2-step partial arc the nozzle ring is divided into two segments. The main air supply is split and led to each of these segments through separate throttling valves. The control is so designed that both segments are used at the minimum energy conditions; the turbine and nozzles are designed the same as in the throttling system. The segments are so split that little throttling is required at cruise load and energy. The machine is operated at cruise with one valve open and the other shut off completely. At this condition the losses result from running the turbine in the inactive portion of the arc. The test program showed these losses to be less than those encountered with the variable-angle nozzle at the same energy condition. As result of this program the 2-step partial arc system was selected for the Turbonator.

The nozzle and turbine wheel design was conventional except for the frangible feature. The turbine buckets were designed to fail on the wheel at a speed above the transient maximum speed reached after operation of the overspeed trip, but below the burst speed of any other part of the rotor. The flange directly over the turbine blading was designed to contain these buckets inside the machine. Thus in the event of a possible double failure of both the basic speed-control system and the overspeed protection, the maximum speed of the shaft is limited by the failure of the buckets but without damage to the aircraft. Tests conducted on more than 20 wheels established that all buckets failed simultaneously on any

Fig. 6. Rotor and electric stator

wheel and all wheels failed between 34,000 and 38,000 rpm. The method is currently qualified for flight.

Bearing Design

The bearing at the turbine end of the machine is a hydrostatic air bearing designed to operate on jet engine bleed air. The advantages for this type of bearing are its extremely low frictional loss and its ability to operate without wear and over a wide temperature range without special lubricants. Air bearings require a relatively large surface area per unit load, but this is compatible with the surface area and diameter of the generator rotor. The bearing was designed to operate with a nominal radial clearance of 0.004 inch with design loads taken from the aircraft application. During the testing of the machine the air bearing was instrumented with capacitance probes which permitted continuous measurements of both running radial clearance and eccentricity. These measurements demonstrated that the unique method of prestressing the journal was effective in limiting the radial growth of the rotor and also that the temperature differential across the bearing was smaller than had been anticipated when running with a hot rotor and cold turbine exhaust. The results indicate that materials with like coefficients of expansion should be used on both journal and bearing rather than a higher coefficient of thermal expansion on the bearing as was done in the test machine. The air bearing has a sufficiently low spring constant to bring the first critical well below running speed. This critical is quite sharply tuned and is readily varied by varying the supply pressure. During runs with a preset amount of unbalance in the rotor, wide excursions could be observed when passing through the critical speed. In the later stages of the testing program the bearing was lined with plastic, permitting the bearing to rub when passing through the critical, and to carry transient loads in excess of the design load without damage to the surfaces. The air bearing design employed has a low damping constant and is not suitable for high-vibration environments with forcing frequencies within the critical range.

The lower bearing on the unit is a typical high-speed high-temperature ball bearing. The lubrication system consists of a slinger mounted on the shaft drawing lubricant from the cup up through the bearing. Drain passages are provided around the bearing and back to the cup. The entire capsule is sealed and is pressurized by the thermal expansion of the enclosed air. The machine has been operated with both standard and tool steel bearings using both inner and outer race riding retainers and with both silicone and silicate ester lubricants. The tests demonstrated that the slinger would flood the bearing if the oil level was too high, but that the system was not sensitive to a low oil level provided the slinger was immersed. Flooding resulted in high bearing temperatures while the proper levels gave a rise of approximately 10 C (degrees centigrade) above the exhaust air temperature. The outer race riding cage gave a lower temperature rise than the inner race riding as would be expected in this type of system. Bearing performance was entirely satisfactory and presented no problems during the test program.

Generator Design

The electrical design of the generator portion of the Turbonator presented several interesting problems. Almost all presently used a-c aircraft generators are 400-cps salient-pole machines with 6 to 12 poles (4,000 to 8,000 rpm). The generation of 400 cps at 24,000 rpm required a 2-pole design. There is little experience available in the area of 2-pole designs for aircraft, but industrial machines of this type are common and most use a round-rotor rather than a salient-pole design. In the round-rotor design the field winding is distributed in a large number of small slots rather than being concentrated in one large slot per pole. Experience in the design of 6- and 8-pole aircraft generators has shown that these machines will generally be lighter if a salient-pole construction is used. Because of this experience, a careful study of the relative merits of round-rotor and salient-pole construction was made for the Turbonator.

Complete designs were prepared for each type of machine and sample rotors were built and tested for strength at high speeds. The salient-pole rotor used a unique construction with shafts bolted on at each end of the rotor body so that the winding of the field coil was unusually simple and straightforward. Even with special provisions for retaining end turns, the support of the field coil proved to be the limiting item in the design. The centrifugal forces at the 30,000-rpm overspeed limited the permissible size of the coil winding so much that the salient-pole design was not competitive in weight. Results of the design comparison showed that the salient-pole construction was at least equivalent to round-rotor in manufacturing complexity. In addition, it required a minimum of 5% more weight for the same electrical performance.

Table II. Generator Characteristics

Characteristics at 400 Cps	Turbonator		Salient Pole	
Rating (kva)	40	22	40	40
Poles	2	4	8	8
Per-unit reactance				
Direct axis:				
Synchronous (X_D)	1.75	1.82	2.25	1.71
Transient ($X_{D'}$)	0.14	0.26	0.41	0.29
Subtransient ($X_{D''}$)	0.11	0.14	0.25	0.18
Quadrature axis:				
Synchronous (X_Q)	1.40	0.88	1.27	0.98
Transient ($X_{Q'}$)	0.46			
Subtransient ($X_{Q''}$)	0.11	0.17	0.28	0.20
Armature leakage (X_A)	0.05	0.08	0.18	0.13
Negative sequence (X_2)	0.11	0.16	0.27	0.19
Time constants (milliseconds)				
Direct-axis open circuit ($T_{Do'}$)	116	132	132	179
Direct-axis short circuit ($T_{D'}$)	10.9	18.9	23.8	30.9
Armature (T_A)	2.1	3.5	4.4	3.5
Ratio (X_D/X_Q)	1.25	2.07	1.77	1.75

The round-rotor type of construction finally adopted was much like a scaled-down version of the 3,600-rpm turbine generators used for central-station 60-cps power generation. Fig. 6 shows the external appearance of the rotor and stator. The similarity to conventional turbine-generator designs had the advantage of making available a large amount of design experience directly applicable to the Turbonator.

The Turbonator generator is quite unconventional in appearance when compared to typical a-c aircraft generators. This leads to a suspicion that electrical characteristics will also be significantly different. This did not prove to be the case. Table II compares reactances and time constants for the Turbonator with several more conventional designs (all salient-pole). While all characteristics are of the same order of magnitude, the Turbonator does have somewhat lower reactances and time constants. This will, in general, permit improved performance in aircraft electric systems. One other point of interest is evident in Table II. While the Turbonator rotor is "round" in the sense that it does not have salient poles, it does exhibit the effect of saliency which is characteristic of most aircraft machines. This is evidenced by the ratio of direct to quadrature axis synchronous reactance. This salient-pole effect is due to the concentration of rotor slots in the quadrature axis. These slots cause a significant increase in the effective air-gap length in this region.

Standard aircraft generator design techniques were used for many of the Turbonator performance calculations, but in some cases it was necessary to modify methods to allow for the 2-pole round-rotor construction. Special equations were used for armature winding end leakage reactance. Particular care was also necessary in analyzing the shape of the air-gap flux-density distribution. The relatively wide rotor slot made it necessary to consider the effect of the slots on the gap flux distribution. Special methods were developed to include the stepped magnetomotive force wave due to the distributed field winding and the variation of air-gap permeance due to rotor slots. Performance calculations made in this manner gave excellent correlation with test data. Over-all accuracy on prediction of excitation requirements was within 3%.

The Turbonator rotor is a solid steel forging instead of the laminated structure used in more typical a-c aircraft generators. The steel rotor body provided all of the necessary amortisseur action, but it was necessary to base the calculation of negative sequence and subtransient impedances on central station design methods. Extrapolation of 60-cps methods to the 400-cps turbonator design presented no problems. Accuracy was entirely satisfactory.

Absence of rotor laminations did introduce the possibility of excessive rotor surface losses due to the flux variations associated with the stator slot openings. The frequency of these flux variations in the Turbonator is approximately double that normally encountered in 400-cps aircraft machines. In order to minimize the rotor surface losses, the stator was built with partially closed slots. The slot opening was only 50% of the air gap length. Careful rotor surface temperature measurements have indicated no excessive heating. The large stator slot overhang necessitated the development of special stator winding techniques, but this resulted in an unusually rugged mechanical construction.

The magnetic characteristics of the rotor steel were fully comparable to those of the motor-grade silicon steel more commonly used for salient-pole rotors. Since maximum stresses were in the radial direction, the rotors were upset-forged to give a grain structure favorable for this stress distribution. Magnetization data were obtained by laboratory tests on sample bars cut from the forgings. Because of the small rotor diameter, it was necessary to use longitudinal samples for these tests. Data obtained on the samples proved to be quite conservative since the test pieces were checked in the cross grain direction. Saturation levels in the Turbonator are significantly lower than could be predicted on the basis of the laboratory magnetization data. Test specimens from future forgings will be selected to give data on characteristics in the transverse direction.

The Turbonator test generator is statically excited. Excitation power for the d-c field is obtained from the a-c output with a combination of current and voltage feedback.[5] This approach makes it possible to locate the exciter at some position remote from the Turbonator. The generator construction is equally compatible with other forms of excitation utilizing rotating exciters. Power for control during fault conditions is obtained by rectifying the output of the 1,600-cps permanent-magnet generator mounted on the main shaft. The same permanent-magnet generator provides a speed sensing signal for turbine control purposes.

Since excitation power is returned to the rotor through brushes and slip rings, an extensive test program was carried on in parallel with the Turbonator development to determine the optimum configuration for the current collection system. Various brush holder and slip ring designs and material combinations were evaluated. The final design uses a special brush holder designed to minimize brush pressure variations. Slip ring diameter was kept small enough to hold surface speeds slightly below the speeds slightly below the speeds normally used for 3,600-rpm central-station generators. It proved advantageous to groove the slip rings spirally to insure more stable and lower brush temperatures. Extensive tests under all types of altitude and temperature conditions show indicated brush life exceeding the 500-hour design objective.

Several general trends became evident during the development of the Turbonator design. A number of factors favor a design with a relatively small diameter-to-length ratio:

1. Rotational stresses make it desirable to keep the rotor diameter low.
2. The large stator yoke section needed for a 2-pole design favors a small diameter since this minimizes the length of path over which the fixed cross section must be maintained.
3. The long end turns required by 2-pole windings make a small-diameter design more favorable since this will maximize the ratio of useful to total copper.
4. Low diameter-to-length ratio in air passages make for more effective generator cooling.

The large rotor and stator iron sections in the 2-pole design also favor a relatively low flux, high reactance design. This leads to a lighter machine because of the relatively large ratio of iron to copper weight in the typical 2-pole designs. Test results showed that the large number of stator slots per pole which can be used in a machine like the Turbonator simplify the problem of eliminating harmonics in the output voltage. Total rms harmonic content was less than 2% and the highest individual harmonic was 1.13% of the fundamental voltage. The relatively low leakage and subtransient reactances permit excellent system performance.

When development of the Turbonator started, it was felt that centrifugal stresses would be a major design limitation. While this was correct for the sample salient-pole rotor, it did not hold true for the round-rotor design. Because the speeds involved were higher than most existing designs, careful mechanical design was required. However, the design stress levels used were not significantly higher

than those used for 6,000- and 8,000-rpm machines. It was not necessary to make any compromises with standard practices regarding factors of safety. All rotor stresses including concentration factors were kept below 50% of the 0.02% yield stress for the material involved, at the 30,000-rpm overspeed point. It has been operated at 40,000 rpm without permanent set.

At two points in the rotor, it was desirable to use nonmagnetic materials for structural purposes. Stainless steel might have been used here, but titanium proved to be much more attractive. The wedges which hold the rotor windings in place are titanium, as are the retaining rings which support the end turn portion of the windings. These retaining rings are required to support a relatively large mass of copper and, consequently, are quite thick. Because of this thickness, a fairly large portion of the ring's stress is due to its own mass. Use of titanium here provides three advantages:

1. Since titanium is appreciably lighter than steel it reduces over-all generator weight.

2. Titanium has better strength characteristics than most nonmagnetic stainless steels.

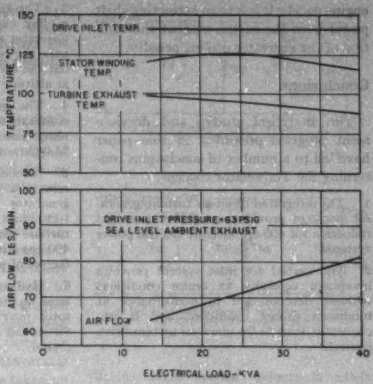
3. The lighter weight of the titanium reduces total stress in the retaining rings by about 25%.

Similar advantages are gained by the use of titanium for rotor slot wedges.

The retaining rings provided support for the end turns on the rotor, but considerable care was necessary to prevent shifting of the coils under the influence of centrifugal and thermal stresses. It was necessary to insure that the coil load was uniformly distributed over the inner surface of the retaining rings. In order to accomplish this the end turns were impregnated with a high-temperature potting compound and wrapped with silicon-impregnated glass roving. The insulated structure was then machined to close tolerances so that the rings could be fitted over them with a uniform shrink fit. This technique stressed the end turn insulation in compression under all speed conditions and effectively prevented any coil motion.

High-temperature insulation was used throughout the generator. Basic insulation structure is glass laminate and mica with a special high-temperature silicon impregnant. Rotor insulation presented some special problems. Compressive stresses on the rotor insulation did not exceed 6,000 pounds per square inch even at overspeed, and this is well within the capabilities of the insulation used. However, care was necessary in the selection of laminated insulation to be used in a radial plane. It might be possible to delaminate some materials under the combined effects of speed and temperature. No problems of this type were experienced.

Fig. 7. Sea level cooling test data



Cooling

Fig. 5 shows the cooling passages in the rotor and stator of the machine. Turbine exhaust air is passed over the stator and lower bearing cartridge. Restrictions due to the stator and stator support tubes in this path provides the static pressure drop required to force cooling air through the gap and through passages in the rotor body. These passages lead from the ends of the rotor to exhaust holes near the center of the stack where the air is exhausted into the gap and through radial slots in the stator to the main exhaust stream. The high surface velocity on the rotor and the pumping action of the air passages, within the rotor contribute to the effectiveness of the rotor cooling. For clarity, Fig. 5 does not slow the tube which connects the boss on the casing downstream of the turbine to the upper end of the rotor.

Sample performance from sea level testing is shown in Fig. 7 where both air flow and temperature are plotted. As an example of the cooling capacity of the unit, it can be seen that with a 33-kva load imposed on the unit and with the unit inlet air bleed condition of 63 psia (pounds per square inch absolute) and 148 C, there is 1.34 pounds per second of 93 C air available at the turbine exhaust for cooling. The stator winding temperature is stabilized at 123 C by this cooling air. The trend of lower cooling air temperatures and increased flow with increasing load is quite evident from Fig. 7.

Typical of the cooling capability demonstrated under altitude operation was a 37,000-foot test point with 207 C, 35-psia inlet conditions and a 57.5-kva load. At this condition 0.89 pound per second of air flow at 98 C was available at the turbine exhaust for cooling. This resulted in a stabilized winding temperature of 180 C.

Close correlation between test winding temperatures at all operating conditions was obtained by using the approximate relationship

$$\Delta T = \frac{K\left(\frac{1}{\eta_g} - 1\right) \text{load}}{W^{0.8}}$$

where

ΔT = winding temperature–turbine exhaust temperature
W = air flow rate
K = a constant which may be determined from any one test point
η_g = generator efficiency

Recently design studies have been made applying the Turbonator to aircraft in the Mach 2.5 to 3.5 speed range and the over-all aircraft penalties associated with the system. At these speeds there is sufficient energy in the ram air to operate the machine and the exhaust temperatures and air flow rates provide comparable winding temperatures to those realized in subsonic aircraft using bleed air. Therefore, if a switching device is provided which will switch the turbine supply air from engine bleed to ram, the existing machine can be operated at aircraft speeds of Mach 2.5 to Mach 3.5 without external cooling or increased temperature limits. The system penalty studies show that the aircraft penalty associated with using this system and exhausting the air to the

engine nacelle is less than extracting shaft power from the engine without the addition of the generator cooling penalty.

Conclusions

The analytical studies and development program presented in this paper have led to a number of conclusions concerning the Turbonator concept.

1. The integrated bleed-air turbine-generator package provides a significant weight reduction for 400-cps aircraft electric power systems.

2. The partial arc inlet system provides maximum economy at cruise conditions without deterioration in performance at minimum energy conditions and is mechanically feasible for bleed-air turbine drive control.

3. The turbine exhaust air has sufficient cooling capacity to maintain satisfactory generator winding and bearing temperatures over the complete range of operation without the necessity of derating the equipment at altitude or ground idle conditions.

4. The generator design adopted gives constants suitable for aircraft generator systems and has proved the feasibility of 24,000-rpm aircraft generators.

5. Although forged round-rotor design represents a departure from current aircraft generator design practice, 60-cps design techniques developed for central station turbine generators can be extrapolated to 400-cps aircraft machines, holding design factors of safety to acceptable limits.

6. Hydrostatic air bearings provide a practical method for supporting a 24,000-rpm rotor provided the external vibration environment is not severe in the critical speed region.

7. Splash-lubricated ball bearings provide a satisfactory bearing system for 24,000-rpm aircraft generator and turbine rotors.

References

1. Direct Engine-Driven Unconventional Electric System, **K. Martinez.** *Paper 228*, Society of Automotive Engineers, New York, N. Y., 1957.

2. Ram-Air Cooling Systems for Aircraft Generators, **Richard M. Moroney.** *AIEE Transactions*, vol. 76, pt. II, Sept. 1957, pp. 217–21.

3. Aircraft General Specification for A-C Air-Cooled Generators and Rrgulators. *Military Specification MIL-G-6099A(ASG)* Mar. 25, 1957.

4. An Electro-Hydraulic Speed and Load Division Control for Constant-Speed Air-Turbine Drives, **P. Dantowitz, L. G. Morris.** *AIEE Transactions*, vol. 77, pt. II, May 1958, pp. 99–106.

5. A Static Exciter for Aircraft A-C Generators, **H. H. Britten, D. L. Piotte.** *Ibid.*, Sept., pp. 271–77.

Random Paralleling of Aircraft Alternators

F. P. de MELLO
MEMBER AIEE

H. M. RUSTEBAKKE
MEMBER AIEE

MANY large aircraft in use today have constant-frequency electric power systems. The alternators in several of these systems can be operated in parallel and the procedure of connecting machines together electrically usually involves synchronizing and paralleling. Here synchronizing means bringing the machines into proper frequency and phase position, while paralleling is electrically connecting the machine armature circuits through a switching device. The terminal voltages will be held within satisfactory limits by the automatic voltage regulators.

The time and equipment used to synchronize[1–3] then involves only two quantities, frequency and phase position. Since the hydraulic constant-speed drives will hold the frequency of each alternator within narrow limits, the matching of phase position is the main function of synchronizing equipment. If the paralleling were to be performed at any phase angle, as in random paralleling, the generating equipment may be subjected to transient torques. Also, the bus voltage would experience a transient that may be severe.

It was the purpose of this study to investigate the magnitude and nature of the transient alternator torques and bus voltages following random paralleling on a typical aircraft power system.

Paper 58-830, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing April 10, 1958.

P. P. de Mello and H. M. Rustebakke are with the General Electric Company, Schenectady, N. Y.

The authors gratefully acknowledge the assistance of J. T. Duane, Leon Klein, M. J. Powell, A. J. Wood, and Dr. R. T. Smith in reviewing the manuscript and making helpful suggestions.

Nomenclature

e = voltage
I, i = current
x = inductance
r = resistance (without subscript means armature resistance)
ψ = flux linkage
ω = instantaneous angular speed
t = time
p = differential operator d/dt
R = load resistance
L = load inductance
x_l = generator leakage inductance
x_{ad} = direct-axis air-gap magnetizing inductance
x_{aq} = quadrature-axis air-gap magnetizing inductance
$f(\psi_{md})$ indicates that the generator saturation effects are included. It was assumed that these effects were a function of ψ_{md}.
δ = angle by which the oncoming alternator rotor lags the system alternator rotor
T_g = alternator air-gap torque

Subscripts

d = direct axis
q = quadrature axis
f = field
k = amortisseur
m = mutual

Method of Solution

The basic circuit equations used for the alternator were those developed by Concordia[4] based on Park's 2-reaction theory. The hydraulic ball-pump drives were represented by their appropriate equations. These equations were solved on an electronic differential analyzer. The aircraft engines were considered to be of infinite inertia in this problem. Two drive circuits and two alternator circuits on the computer were sufficient to represent a 2- or a 4-alternator system; see Fig. 1. The constants of alternator 1 in Fig. 1 could be altered to represent three machines as well as one machine without changing the circuitry. Solutions were determined for paralleling one loaded alternator with another loaded alternator. Further, solutions were made for paralleling one alternator, either loaded or unloaded, with three loaded alternators. In each case the effect of feeder and static exciter impedance between alternator and bus was considered; i.e., runs were made with and without this impedance. Machine constants representative of a typical installation were used and are listed in Appendix I.

Pessimistic approaches were used in the solution of torques and transient

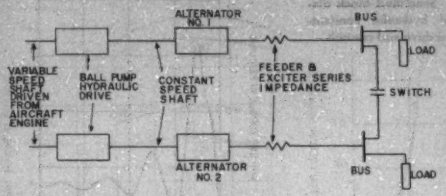


Fig. 1. Aircraft 400-cps electric power system. Alternator 1, which may represent one or three alternators, is considered as the operating system. Alternator 2 is considered as the machine to be paralleled with the system

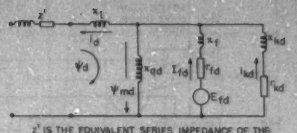


Fig. 2 (right). Approximate direct-axis equivalent circuit for an alternator

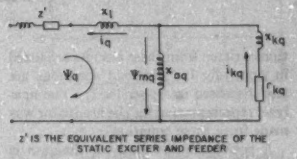


Fig. 3 (below right). Approximate quadrature-axis equivalent circuit for an alternator

voltage. The machine was considered to be at −55 C (degrees centigrade) in the determination of torques. This gave the longest time constant for the d-c armature transient resulting in maximum transient torques. Also, as stated previously, the impedance between alternators was removed for some runs. The resulting torques should be greater than those experienced in random paralleling in aircraft installations using alternators of comparable design.

The determination of transient voltage, was made with the machine fields at 250 C and with impedance between alternators. Under this temperature condition, the ability of the exciter to restore normal voltage is less than at lower temperatures.

In order to keep the scope of this study within practical bounds, the extent to which different conditions were studied had to be limited. For a particular load condition and speed difference there is a switching angle that is more unfavorable than others. Several runs were taken on either side of this angle, but no attempt was made to try the angles where little disturbance would occur. Also as mentioned before, extreme temperature conditions were used. Thus an idea was obtained of the maximum torques and the poorest transient voltage performance that could be expected from the practice of random paralleling on this system.

The assumptions made in this study included:

1. Only simultaneous 3-pole closing is considered.
2. A single mutual inductance exists among the three circuits in the direct axis: field, armature, and damper.
3. Saturation occurs principally along the main pole axis.
4. The torques imposed on the main aircraft engine drive shaft (ball-pump-drive input) do not affect its rotational velocity.

Analyzer Representation for the Alternator

The direct-axis equivalent circuit of an alternator[8] is shown in Fig. 2 and that for the quadrature axis in Fig. 3. The equa-

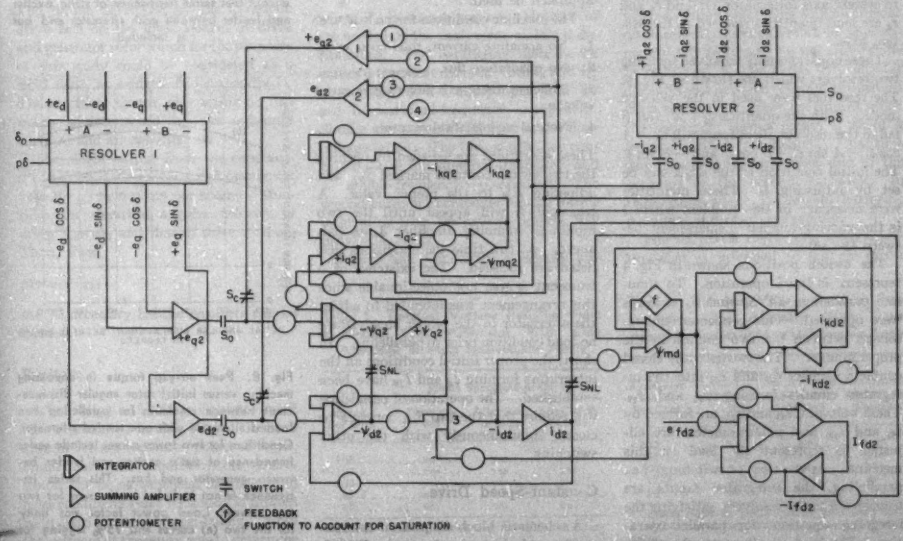


Fig. 4. Electronic differential analyzer circuit used to solve equations for alternator 2 of Fig. 1. Resolvers and paralleling switches are also shown

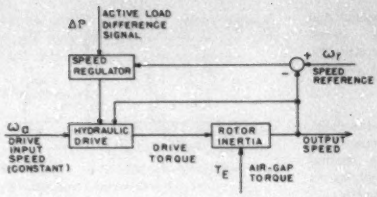


Fig. 5. Simplified block diagram of hydraulic constant-speed drive and controls

tions written from these circuits are listed in Appendix I. The load equations are also discussed in Appendix I. The analyzer representation for the alternator was made from these equations.

A portion of the analyzer diagram is shown in Fig. 4. This analyzer circuit is for alternator 2 of Fig. 1 and solves the alternator equations. A similar circuit would describe the equations for alternator 1. The connection between alternator circuits must satisfy the following equations:

$$i_{d1} = i_{d2} \cos \delta + i_{q2} \sin \delta + i_{d1} \quad (1)$$

$$i_{q} = i_{q2} \cos \delta - i_{d2} \sin \delta + i_{q1} \quad (2)$$

$$e_{d2} = e_{d} \cos \delta - e_{q} \sin \delta \quad (3)$$

$$e_{q2} = e_{q} \cos \delta + e_{d} \sin \delta \quad (4)$$

where subscript 1 is a system quantity (alternator 1); subscript 2 is ar alternator 2 quantity (oncoming alternator); i_d and i_q are load quantities taken with respect to system axis (alternator 1); and e_d and e_q are load quantities referred to the system axis.

These equations may be solved by using two resolvers which are shown in Fig. 4. The resolvers were used in the velocity mode. With the quantities $p\delta$, A, and B fed to the resolver, its outputs will be $-A \cos \delta$, $-A \sin \delta$, $+B \cos \delta$, and $-B \sin \delta$. The initial condition of the angle can be set by adjusting δ_0. These quantities were necessary in this problem where δ is the varying angular displacement between the rotors.

The switch positions shown in Fig. 4 represent isolated operation. To simulate paralleling, all S_0 and S_c switches were operated. This connected the resolvers between the two circuits in the proper manner. The two normally closed contacts connect e_{d2} and e_{q2} into the integrator circuits forming ψ_{d2} and ψ_{q2}. These voltages, e_{d2} and e_{q2}, are formed by i_{d2} and i_{q2}, and potentiometers are adjusted to represent the load on this machine. At the time of switching, i.e., paralleling, the integrator inputs are transferred to the resolvers, satisfying the foregoing equations for parallel operation. The load on alternator 2 appears lost at this point. However, alternator 1 (not shown in the diagram) simultaneously experiences a change in load equal to that lost by alternator 2. Thus the conditions before and after paralleling are satisfied.

Representation for No-Load Condition

The load was represented as a series resistance and inductance. Thus to represent a short circuit on the alternator, it was only necessary to set potentiometers 1, 2, 3, and 4 of Fig. 4 to zero, or a short circuit on the alternator is represented when the inputs, hence the outputs, of summing amplifiers 1 and 2 are zero.

The representation for a no-load condition on the alternator requires that the input (and output) of amplifiers 1 and 2 be infinite. This demands that another approach be used.

The machine conditions for no load are:

1. No armature current, thus no i_d or i_q.
2. No quadrature flux.
3. Sufficient direct-axis flux to give rated voltage.
4. Normal no-load field current.

These conditions are satisfied by opening the two switch contacts marked S_{NL} and adjusting E_{fd} to the proper value. A transient i_d will appear until the two inputs to summing amplifier 3 are balanced. At this time the conditions just listed are satisfied. The existence of a transient i_d was not objectionable since this arrangement was intended to adjust the alternator to the proper steady-state no-load condition prior to paralleling. In short, the proper initial conditions on the integrators forming ψ_d and I_{fd} have been established. The operation of paralleling will require that the two S_{NL} contacts be closed simultaneously with the other switching.

Constant-Speed Drive

A schematic block diagram of the constant-speed drive is shown in Fig. 5.

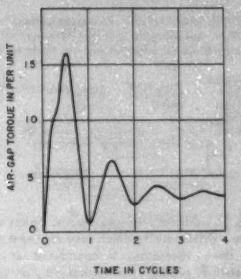


Fig. 6. Air-gap torque in oncoming machine versus time for paralleling one unloaded alternator with loaded 3-machine system. Paralleling switch was closed at 120 degrees between rotors. No external impedance was connected between machines and bus

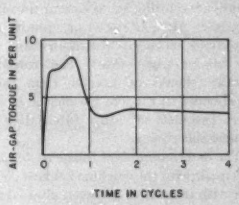


Fig. 7. Conditions are same as for Fig. 6 except that series impedance of static exciter and feeder between each alternator and bus is included

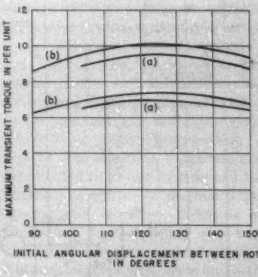


Fig. 8. Peak air-gap torque in oncoming machine versus initial rotor angular displacement between machines for paralleling one loaded alternator with one loaded alternator. Conditions for two lower curves include series impedance of static exciter and feeder between generator and bus. This series impedance is not included in the runs for two top curves. Load power factor was unity for the two (a) curves and 75% lagging for the two (b) curves

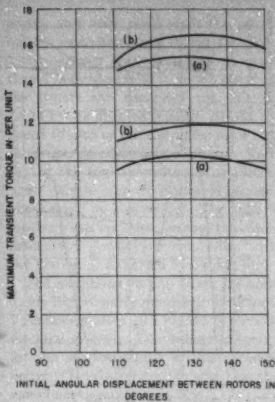


Fig. 9. Same as Fig. 8 except for paralleling one loaded alternator with three loaded machines

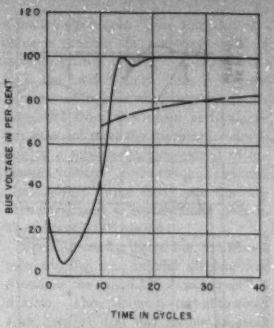


Fig. 10. Bus voltage versus time for paralleling unloaded alternator with loaded one at 180 degrees between rotors. Machines were running at same speed

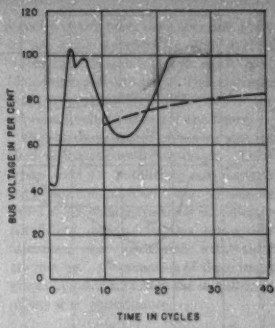


Fig. 11. Bus voltage versus time for paralleling one loaded alternator with three loaded machines at 210 degrees. Single machine was running 5% faster than others

There are four inputs to the drive circuit. They are reference speed ω_r, drive input speed ω_0, air-gap torque T_g, and active load difference signal ΔP. The drive is essentially a zero steady-state error speed regulator with a bias signal from the real load division circuit. The drive input speed which is the aircraft engine speed was assumed constant in this problem. The torque developed by the hydraulic drive acts on the mass system of drive and generator rotor which for the purposes of this study could be considered as a rigid mass, as explained in Appendix II. The electrical torque, also acting on this mass, is determined by the armature currents and air-gap flux.

The single output from the constant-speed drive is alternator speed or angular velocity ω_L in radians per second. Consider the electrical angular velocity of drive 1 as ω_{L1} and that of drive 2 as ω_{L2}. Then, since

$$p\delta = \omega_{L1} - \omega_{L2} \tag{5}$$

the $p\delta$ necessary for the resolvers can be formed.

Exciters

The exciters and voltage regulators were also represented in this study. From a system-voltage-recovery point of view, under hot field conditions the exciter would not be as beneficial as under cold field conditions since its action is limited by a fixed exciter-voltage ceiling. Since the purpose of the study was to determine the limitations of random paralleling, the pessimistic case of a 250 C field was assumed for transient-voltage tests. A field temperature of −55 C was used in the determination of maximum transient torque.

Torques

The expression for torque in a synchronous machine is

$$T_g = i_q\psi_d - i_d\psi_q \tag{6}$$

where T_g is air-gap torque. Multipliers are needed on the analyzer to satisfy these equations. The paralleling disturbance causes a transient current (both i_d and i_q) to contain a d-c component and an a-c fundamental-frequency component. Similar components will exist in the flux-linkage terms. The torque will then contain a d-c, a fundamental-frequency, and a double-frequency component. The second-harmonic component in the torque originating from the saliency effects, will usually be small.

Figs. 6 and 7 show the alternator air-gap torque when it is paralleled at no load with a 3-generator system. The initial rotor angular displacement is 120 degrees for either trace. Fig. 6 shows the torque with no impedance between alternators. Fig. 7 shows the torque with the series impedance of the static exciters and feeders included. The series impedance not only lowered the peak torque but also caused a much more rapid decay of the transient.

Examination of many torque traces confirms the fact that maximum torque will result from paralleling with an angular displacement between rotors of approximately 120 degrees.[6] Figs. 8 and 9 show how the peak torque varies with angular displacement between rotors at the instant of paralleling. Further, the time at which the maximum torque occurs is about a half-cycle after paralleling. This time was increased in some cases due to the presence of the second harmonic. In those cases where the second harmonic is evident, it decays very rapidly.

The alternator shafts have a shear section that is intended to fail on excess

Table I. Transient Voltage Resulting From Paralleling One Alternator With One Alternator at Most Unfavorable Angle

Alternator Speed, Per Cent	Meets Specification Curve		
	Through-out	After 0.05 Second	After 0.075 Second
105	no*	yes	yes
102.5	yes	yes	yes
100	yes	yes	yes
97.5	no	no	no
95	no	no	no

* Outside curve for paralleling an unloaded alternator.

Table II. Transient Voltage Resulting From Paralleling One Alternator With Three-Alternator System at Most Unfavorable Angle

Alternator Speed, Per Cent	Meets Specification Curve		
	Through-out	After 0.05 Second	After 0.075 Second
105	no	no*	no*
102.5	yes	yes	yes
100	yes	yes	yes
97.5	no†	no†	no*
95	no	no	no

* Outside curve for paralleling a loaded alternator.
† Outside curve for paralleling an unloaded alternator.

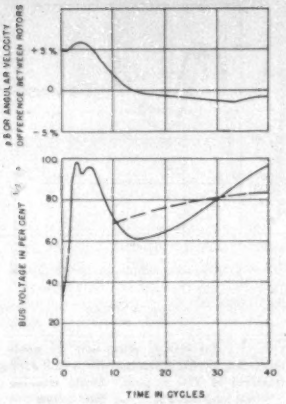


Fig. 12. Bus voltage and angular velocity difference versus time for paralleling one loaded alternator with three loaded alternators. Initial angular difference was 180 degrees and single unit was at 5% overspeed

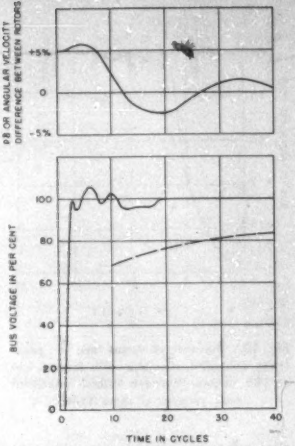


Fig. 13. Conditions same as for Fig. 12 except that machine is paralleled with single loaded alternator

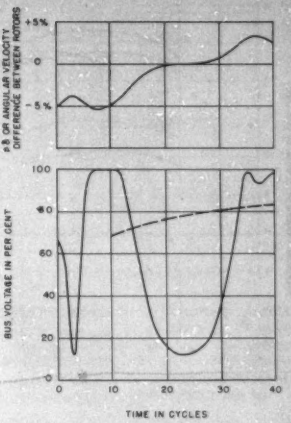


Fig. 14. Bus voltage and angular velocity difference versus time for paralleling one unloaded alternator with one loaded alternator. Unloaded machine is paralleled when running 5% slower than other unit and its rotor lags loaded machine rotor by 120 degrees

torques before more serious damage can occur. If the peak transient torques of Fig. 9 were constant or of a much lower frequency, they would be imposed on this shear section and failure could result. However, the rotor inertia greatly attenuates this torque at the frequency involved and the actual shaft torque is quite small. This phenomenon is explained in some detail in Appendix II.

In addition to being of high frequency, i.e., 400 cps (cycles per second), the transient torques decay very rapidly as shown in Figs. 6 and 7. Furthermore, normal load torques are restored quite rapidly as the rotors move into their normal operating position.

The air-gap torque can be considered to act directly on the stator core pins and the frame mounting bolts. These torques are small compared with the shear strength of these pins and bolts.

Transient Voltage

The transient-voltage curves of Fig. 1 in specification *MIL-E-7894A* (Aircraft Standard Group)[7] were used as a guide or benchmark in the evaluation of voltage-recovery performance. Some of the transient-voltage curves fell outside the specification curves by only a small amount, see Figs. 10 and 11, and therefore two other measurement standards were established. One was taken as requiring that the voltage stay within the envelope shown in the specification beyond 0.05 second (20 cycles). The other would be that it stay within the envelope beyond 0.075 second (30 cycles). Tables I and II show a tabulation of the results.

The results of these tables apply to paralleling at the most unfavorable angle. It is evident that for any given condition of speed difference and load there is a large region of angles for which paralleling would cause a very minor transient-voltage disturbance.

The most unfavorable angle would be that for which the relative speed difference between rotors, following paralleling, is reduced to zero as the rotors move to phase opposition. This critical angle of the oncoming rotor with respect to the system rotor is a function of the relative speed difference between machines. For the speed differences considered, this angle lies in the second or third quadrants, depending on whether the oncoming machine is fast or slow, respectively.

An examination of Table I indicates that good transient-voltage performance can most easily be met by driving the lightly loaded alternator slightly fast. The tests indicate $2^1/_2\%$ fast (10 cps) to be a good value. The difference between having a lightly loaded oncoming machine run slightly fast rather than slow with respect to the loaded system can be explained in this manner. Upon switching at large angular differences between rotors, the drop in voltage causes a drop in system load which tends to accelerate the loaded system. With the oncoming machine running fast, this effect is in the direction to decrease the relative speed difference between machines and thus make it easier for the machines to pull into step. If the oncoming machine were running slow, the drop in system load would tend to increase the relative speed between machines and make pulling into step more difficult.

Table II shows that random paralleling of one alternator with a system of three alternators will cause transient-voltage disturbance somewhat more severe than for the case of paralleling against a single alternator system. The initial voltage dip is of course less than with a single-alternator system since the bus is no longer at the impedance center as is true for the case of paralleling two equal-size machines. The difficulty experienced by a 3-unit system in meeting the specified requirements is not in magnitude of initial drop but in the duration of the reduced voltage during the rotor oscillations. Fig. 12 shows an example of low voltage for a long duration. The $p\delta$ curve shown in Fig. 12 gives an explanation of why the longer period of low voltage exists. The paralleling took place with the rotors 180 degrees out of phase and the oncoming alternator about 5% overspeed. Under these conditions, the voltage re-

covery was very rapid since the rotors quickly came into phase (about 4 cycles). The alternator rotor was then at more than 5% overspeed so it pulled ahead of the system and depressed the voltage. At 14 cycles after paralleling, $p\delta$ was reduced to zero or the alternator rotor was the proper speed but out of phase. Note that at this time the voltage is at a minimum which also indicates maximum phase angle on this swing. The voltage recovery was very slow because the alternator rotor slowly came into phase with the system as indicated by a small $p\delta$.

The conditions for Fig. 13 are the same as for Fig. 12 except that Fig. 13 is for a 1-machine system and Fig. 12 is for a 3-machine system. The voltage recovery for the single-machine system is much better. It is interesting to note that the maximum negative $p\delta$ in Fig. 13 is twice that of Fig. 12. The system inertia is the significant difference between the two cases. The lighter inertia system will respond to the "synchronizing" torques by greater velocity. Thus, is this case, it will move more quickly toward the other rotor giving a greater $p\delta$ (see equation 5).

Fig. 14 shows the speed difference curve ($p\delta$) and the transient bus voltage when one unloaded alternator is paralleled with a loaded alternator. The unloaded machine is running 5% underspeed and is paralleled with its rotor 120 degrees behind the loaded unit. This is a good example of a case that appears to be successful from the initial performance but does not meet the specification curve due to a subsequent long period of depressed voltage. Even though the words "long period" were used, note that the entire transient is substantially over in 34 cycles or 0.085 second.

The first voltage dip of Fig. 14 shows a pole slipping, i.e., angular displacement goes from the initial 120 degrees past 180 degrees. Thus, the rotor started to accelerate (see $p\delta$) but was unable to avoid slipping. At 180 degrees the bus voltage appears to be about 12%, not zero. This is because one machine was loaded prior to paralleling and therefore had a greater air-gap flux (or generated voltage) than the unloaded alternator.

After the alternator rotors passed through the 180-degree point, the forces pulling the rotors together further reduce the speed of the slow rotor thereby increasing the relative slip between machines. This causes a very pronounced second swing. If this swing velocity were great enough, the machines would go through additional slip cycles. With the velocity shown here, the rotors pull apart to an angle approaching 180 degrees where the rotor speeds become equal ($p\delta = 0$) and then pull into step. The relatively long time at a large angle causes the depressed voltage shown in the figure.

Fig. 15 shows traces of bus voltage for three different paralleling angles. These traces were for paralleling one machine with only one machine. Fig. 16 shows a series of bus voltage traces for six different paralleling angles. In this series, one alternator was paralleled with three alternators. Examination of these traces gives a good idea of how the initial angle affects the performance.

Conclusions

The torques encountered on random paralleling under normal conditions are not destructive to the equipment considered in this study. The high-transient air-gap torques are of fundamental frequency that decay very rapidly. The transient torques transmitted through the shaft are much smaller due to the large rotor inertia relative to the drive and also to the fact that the natural frequency of the drive rotor 2-mass system is not in the vicinity of 400 cps. Paralleling torques are of short duration that change into nor-

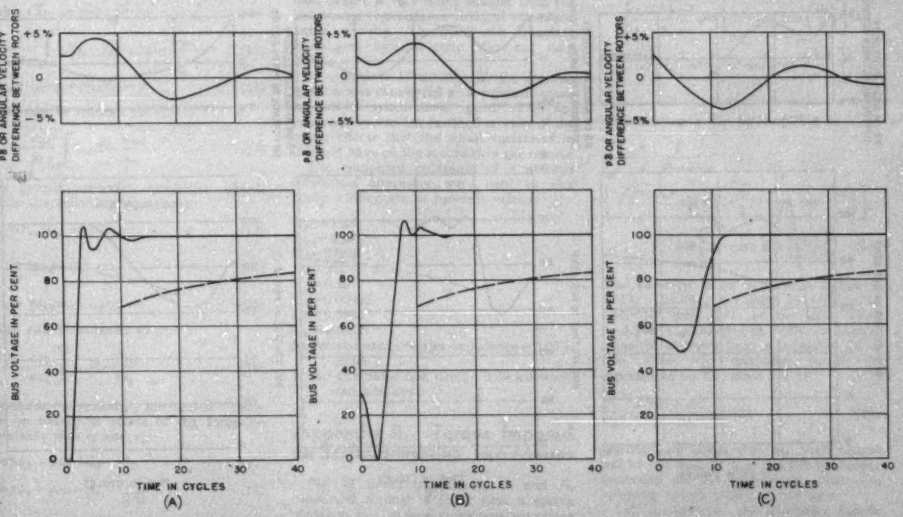


Fig. 15. Bus voltage and angular velocity difference versus time for paralleling one loaded alternator with one loaded alternator when speed difference is 2½%. Initial rotor angular difference is:

A—180 degrees B—210 degrees C—240 degrees

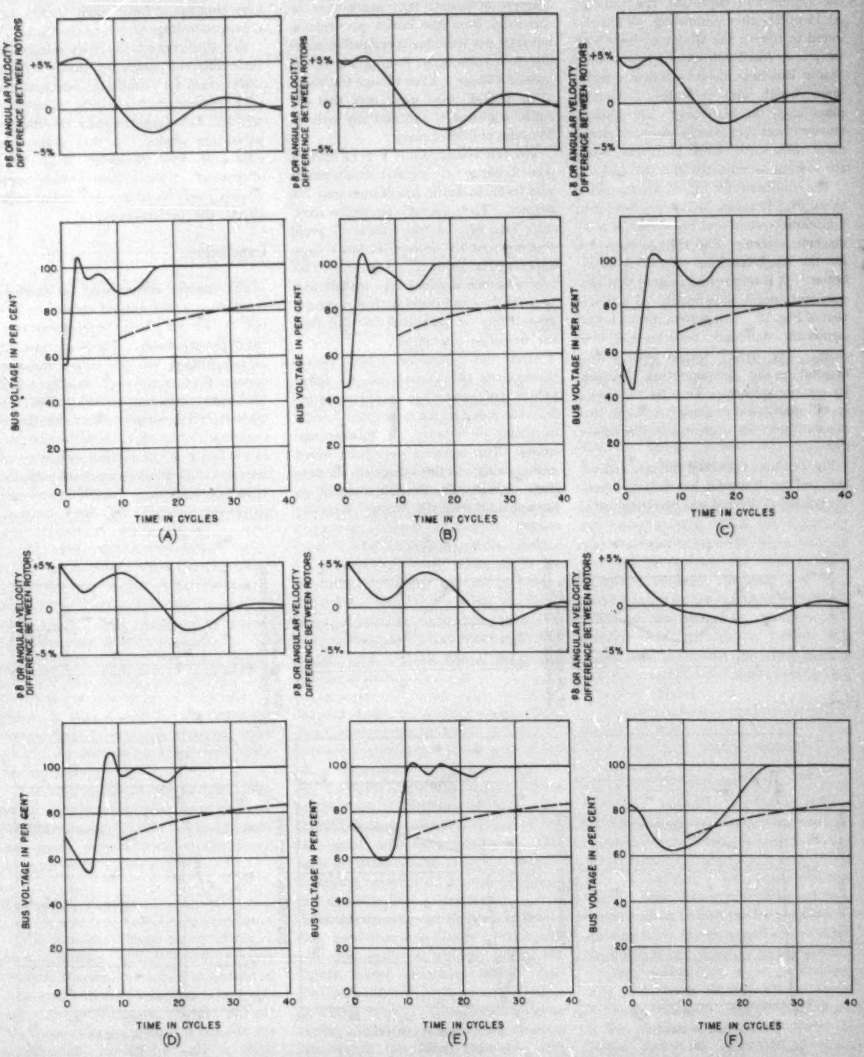


Fig. 16. Bus voltage and angular velocity difference versus time for paralleling one unloaded alternator with three loaded alternators. Unloaded machine is running 5% overspeed. Initial rotor angular difference is:

A—180 degrees B—210 degrees C—240 degrees D—270 degrees E—285 degrees F—300 degrees

mal load torques as the rotors come into phase.

For the system considered in this study, random paralleling appears quite feasible.

Certain guides can be recommended for random paralleling with the least voltage disturbance for systems similar to the one studied in this paper. One recommendation when paralleling an alternator with another is to run the lightly loaded alternator slightly fast. When paralleling a single machine with a loaded multimachine system, it is better to avoid underspeed when the single unit is unloaded.

Appendix I. The Alternator and Load Equations

The direct-axis equivalent circuit of an alternator[3] is shown in Fig. 2 and that for the quadrature axis in Fig. 3. The necessary equations may be written from these circuits. For the direct axis the following equations may be written:

$$\psi_{md}=x_{ad}(I_{fd}+i_{kd}-i_d)-f(\psi_{md}) \tag{7}$$

where the $f(\psi_{md})$ term corrects for the effect of saturation.

$$i_d=\frac{1}{x_l}(\psi_{md}-\psi_d) \tag{8}$$

$$d\psi_{md}+pI_{fd}x_f+I_{fd}r_{fd}-E_{fd}=0 \tag{9}$$

which can be reduced to

$$I_{fd}=\frac{1}{x_f}\int(E_{fd}-I_{fd}r_{fd})dt-\frac{\psi_{md}}{x_f} \tag{10}$$

$$p\psi_{md}+i_{kd}r_{kd}+pi_{kd}x_{kd}=0 \tag{11}$$

which can be reduced to

$$i_{kd}=-\frac{r_{kd}}{x_{kd}}\int i_{kd}dt-\frac{\psi_{md}}{x_{kd}} \tag{12}$$

The quadrature-axis equivalent circuit yields the following equations:

$$\psi_{mq}=x_{aq}(i_{kq}-i_q) \tag{13}$$

$$i_q=\frac{1}{x_l}(\psi_{mq}-\psi_q) \tag{14}$$

$$p\psi_{mq}+pi_{kq}x_{kq}+i_{kq}r_{kq}=0 \tag{15}$$

which can be reduced to

$$i_{kq}=-\frac{r_{kq}}{x_{kq}}\int i_{kq}dt-\frac{\psi_{mq}}{x_{kq}} \tag{16}$$

Equations for ψ_d and ψ_q are needed. They can be found[4] in terms of the foregoing quantities plus e_d and e_q.

$$e_d=p\psi_d-\psi_q\omega-ri_d \tag{17}$$

$$e_q=p\psi_q+\psi_d\omega-ri_q \tag{18}$$

or

$$\psi_d=\int(e_d+ri_d+\psi_q\omega)dt \tag{19}$$

and

$$\psi_q=\int(e_q+ri_q-\psi_d\omega)dt \tag{20}$$

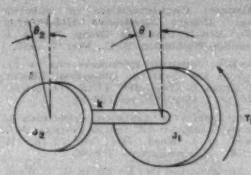


Fig. 17. Torque acting on 2-mass system

External series impedance, such as feeder impedance, can be included by modifying the x_l and r terms in the preceding equations.

The system load was considered to be a resistor R in series with an inductance L. The load current and voltage can be expressed in terms of the direct and quadrature axis of the alternator. If the load is connected to the machine terminals, e_d of the alternator is equal to e_d of the load, etc. The relationships between load voltages and currents are:

$$e_d=Ri_d-Li_q\omega+Lpi_d \tag{21}$$

$$e_q=Ri_q+Li_q\omega+Lpi_q \tag{22}$$

The last term of each equation was neglected. This assumption introduced no error in the case of unity power factor where L is zero. For normal load power factors other than unity, only a slight error was introduced due to the very rapid decay of the transient. Restated, the inductance-to-resistance ratio (time constant) of the load circuit is very much smaller than the inductance-to-resistance ratio of the alternator circuits. Therefore, the transient introduced by this term dies out very rapidly.

In equations 17 through 20 the angular speed ω was considered as a constant equal to normal synchronous speed. This approximation was felt justified in view of the minor effects that the small variations in ω would have on the accuracy of the results.

The computed constants of a 400-cps 8,000-rpm alternator were used in this study. They are in per-unit values:

$X_{ad}=1.35$
$X_{aq}=0.829$
$X_l=0.084$
$r=0.031$
$X_f=0.19$
$r_{fd}=0.00428$
$r_{kd}=0.0808$
$x_{kd}=0.0828$
feeder and exciter series impedance $=0.05+j0.04$
H for alternator and drive $=0.39$ kilowatt-seconds/kva

Appendix II. Torque Imposed on Shaft Connecting Two Masses

Fig. 17 shows two masses, J_1 and J_2, connected through a shaft with a spring constant k. Laplace transform equations can be written for the motion of this system under the influence of the torque T_1.

$$J_1s^2\theta_1(s)=T_1(s)+k[\theta_2(s)-\theta_1(s)] \tag{23}$$

$$J_2s^2\theta_2(s)=k[\theta_1(s)-\theta_2(s)] \tag{24}$$

and shaft torque T is

$$T(s)=k[\theta_2(s)-\theta_1(s)] \tag{25}$$

Let

$$\theta_1-\theta_2=\theta \tag{26}$$

Then from equations 23, 24, and 26

$$\theta(s)=-\frac{T_1(s)/J_1}{s^2+k\dfrac{J_1+J_2}{J_1J_2}} \tag{27}$$

Consider the case of a sinusoidal torque imposed on the mass. Then

$$T_1=T_m\sin(\omega t+\psi) \tag{28}$$

$$T_1(s)=T_m\frac{\omega}{\sqrt{a_0^2+\omega^2}}\frac{s+a_0}{s^2+\omega^2} \tag{29}$$

where

$$a_0=\omega\cot\psi \tag{30}$$

From equations 25, 26, 27, and 29

$$T(s)=k\theta(s)=-\frac{k}{J_1}T_m\sin\psi\frac{s+a_0}{s^2+\omega^2}\frac{1}{s^2+\beta^2} \tag{31}$$

where

$$\beta^2=k\frac{J_1+J_2}{J_1J_2}$$

The inverse transform of equation 31 gives the shaft torque as a function of time.

$$T=-\frac{k}{J_1}\frac{T_m\sin\psi}{(\omega^2-\beta^2)}\times\left\{\frac{\sqrt{\beta^2+a_0^2}}{\beta}\cos\left[\beta t-\arctan\left(\frac{a_0}{\beta}\right)\right]-\frac{\sqrt{\omega^2+a_0^2}}{\omega}\cos\left[\omega t-\arctan\left(\frac{a_0}{\omega}\right)\right]\right\} \tag{32}$$

This expression can be simplified to

$$\frac{T}{T_m}=\frac{k}{J_1}\frac{1}{\beta^2-\omega^2}\times\left\{\frac{\sqrt{\beta^2+a_0^2}}{\beta}\sin\psi\cos\left[\beta t-\arctan\left(\frac{a_0}{\beta}\right)\right]-\cos\left[\omega t-\arctan\left(\frac{a_0}{\omega}\right)\right]\right\} \tag{33}$$

The value of ω for a 400-cps torque is 2,513. Consider the ratio of k/J_1 to be 1.48×10^8 and J_1/J_2 to be 15.5. Then $\beta=4{,}948$ radians/second. The maximum value of the brackets of equation 33 is two. Then the maximum torque T' experienced by the shaft will be

$$\frac{T'}{T_m}=\frac{2.968\times10^8}{\beta^2-\omega^2}=0.163 \tag{34}$$

Equation 34 shows that the shaft torque will be less than 17% of the 400-cps torque impressed on the large mass, when the foregoing system constants are used.

Let J_1 represent the alternator inertia and J_2 the drive inertia. Further, consider that the machine is carrying a load which requires a shaft torque of 1.0 per unit. Assume that the peak transient 400-cps torque at the rotor air gap is 15 per unit. Then the maximum torque imposed upon

the shaft is $1.0+0.163(15)=3.45$ per unit.

Equation 33 shows the shaft torque to be proportional to $1/(\beta^2-\omega^2)$. ω is fixed by the line frequency or 400 cps in aircraft systems. Thus $\omega=2\pi f=2{,}513$ radians/second, β is determined by the "natural" frequency of the system. It is quite apparent that if the system constants were altered so that this natural frequency were to approach 400 cps, the shaft torque would increase.

Equation 25 shows the relationship between angular deflection between shaft ends and transmitted torque. For the machine considered here, the shaft deflection at rated input to the alternator is less than 0.1 of 1 degree. For deflections of this small magnitude the relative motion of J_1 and J_2 can be neglected when computing torques external to this 2-mass system. Thus the two inertias can be combined into a single inertia for such problems.

References

1. Automatic Synchronizing for Aircraft Alternators, **Fred B. McCarty, G. W. Hills.** *AIEE Transactions*, vol. 71, pt. II, 1952 (Jan. 1953 section), pp. 454–58.

2. Considerations Applicable to Automatic Paralleling of Aircraft A-C Generators, **M. J. Powell, E. W. Giloy.** *Ibid.*, vol. 73, pt. II, Nov. 1954, pp. 258–63.

3. Parallel Operation of Airplane Alternators, **D. W. Exner.** *Ibid.*, vol. 62, Dec. 1943, pp. 755–60.

4. Synchronous Machines, Theory and Performance (book), **C. Concordia.** John Wiley & Sons, Inc., New York, N. Y., 1951.

5. An Introduction to Power System Analysis (book), **F. S. Rothe.** *Ibid.*, 1953.

6. Maximum Short Circuit and Faulty Synchronizing Torques on Generator Foundations, **Vernon W. Ruskin.** *AIEE Transactions*, vol. 74, pt. III, Aug., 1955, pp. 618–24.

7. Aircraft Characteristics of Electric Power. *Military Specification MIL-E-7894A (ASG)*, Aircraft Standards Group, Bureau of Aeronautics, Washington, D. C., May 17, 1955.

8. Analog Computer Representation of Alternators for Parallel Operation, **Luis V. Boffi, Vinton B. Haas, Jr.** *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 153–58.

9. The Effects of Abnormal Conditions on Aircraft, Parallel A-C Power Systems, **S. C. Caldwell, A. J. Wood.** *Ibid.*, vol. 72, pt. II, 1953 (Jan. 1954 section), pp. 379–86.

10. Synchronizing Out of Phase, **A. J. Wood.** *Ibid.*, vol. 76, pt. III, Apr. 1957, pp. 1–10.

11. Analogue Computer Representations of Synchronous Generators in Voltage-Regulation Studies, **M. Riaz.** *Ibid.*, vol. 75, pt. III, Dec. 1956, pp. 1178–84.

12. *Discussion* by **C. H. Thomas** of reference 11, pp. 1182–84.

Discussion

A. J. Wood (General Electric Company, Schenectady, N. Y.): The authors have presented an interesting and worth-while paper on the random paralleling of alternators. Although the reported study was concerned specifically with small aircraft machines, it appears that some of their results pertaining to synchronizing torques may well be generalized to include larger machines.

Previous studies of torques due to synchronizing out of phase have neglected the angular speed difference between machines. It would be interesting to know if the authors have available data showing the variation of peak transient electrical torque as a function of the speed difference between the oncoming machine and the system with the synchronizing angle as a parameter. The data presented in Figs. 6 and 7, as well as the data showing $p\delta$ as a function of time, would indicate that the peak transient electrical torque does not depend appreciably on the speed difference, but is primarily determined by the switching angle.

Since the authors have included the effect of voltage-regulator action in their study, it would also be of interest to know if the voltage-regulator characteristics had an appreciable effect on the peak transient electrical and mechanical torques.

F. P. de Mello and **H. M. Rustebakke**: We agree with Mr. Wood that the method of solution in the paper may well be applied to the analysis of synchronizing torques in larger machines.

Regarding the effect of angular speed difference on the peak transient torques, this effect was found to be insignificant for the small angular speed differences considered. The peak transient torque is made up of a large fundamental-frequency component superposed on an aperiodic synchronizing torque component. The fundamental-frequency component is caused by the d-c offset current in the armature whose magnitude is mainly a function of the rotor angle at the instant paralleling occurred.

The synchronizing component of torque is a function of the angular displacement between rotors and it changes as this angular displacement clianges. Since, however, the peak transient torque occurs within half a cycle after paralleling, this angular displacement is not materially affected by small changes in speed; thus peak transient torques are not significantly different for the small changes in $p\delta$.

Again, since the peak transient torques occur within only half a cycle after paralleling, the effects of voltage-regulator action upon these torques is entirely insignificant. This is because of the time constants governing changes in flux due to changes in excitation.

Progress in the Thermal Design of Oil-Cooled Rotating Electric Machines

P. B. RICHARDS
NONMEMBER AIEE

IN THE PAST electrical designers have been hindered by the fact that internal temperatures in a machine were not known with sufficient accuracy. At best they had to be content with average temperatures estimated from measured coil resistances, for example, but the actual location of hot spots and the temperature distribution in the machine components have been difficult to establish. With the increased availability of high-speed digital computers, it is now possible to determine these temperatures. The purpose of this paper is to illustrate how temperatures can be determined in components of electric machines by an iteration technique that has been successfully programmed for the IBM (International Business Machines Corporation) *650* digital computer.

Nomenclature

T = temperature, F (degrees Fahrenheit)
Q = heat generated per unit volume, Btu/(hour)(cubic foot)
K = thermal conductivity, Btu/(hour)(feet)(F)
h = heat-transfer coefficient, Btu/(hour)(square foot)(F)
x,y = rectangular co-ordinates
r = radial co-ordinate

SUBSCRIPTS

s = surface
f = fluid
c = coil

Statement of Problem

Consider the plane cross section of an alternator rotor, normal to the shaft (Fig. 1). Because of symmetry, only the sector shown in Fig. 2 need be studied to determine the temperature distribu-

Paper 58-790, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing April 7, 1958.

P. B. Richards is with Thompson Products, Inc., Cleveland, Ohio.

The author wishes to express his appreciation to John B. Annable, of the Jack & Heintz, Inc. Research Department, who wrote the computer program and obtained the solutions on the IBM *650* digital computer. This program has been submitted to the International Business Machines Program Library, World Headquarters, New York, N. Y., and can be obtained on loan from that organization.

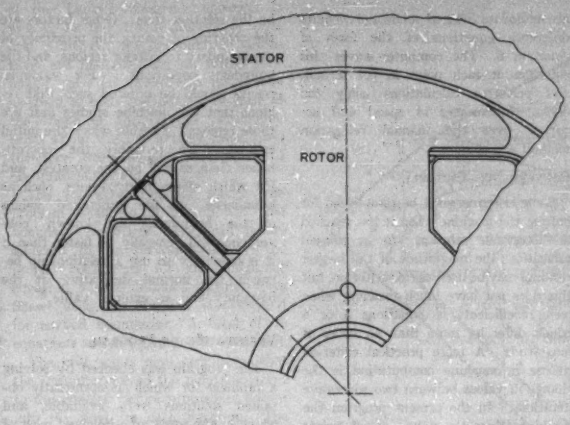


Fig. 1. Cross section of alternator rotor

tion throughout the rotor. The mathematical problem consists in solving the steady-state heat conduction equation throughout this sector, subject to the specified boundary conditions. In the coil, heat is being generated, so the equation to be solved there is Poisson's equation,

$$\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{Q}{K}=0 \qquad (1)$$

while in the rest of the sector Laplace's equation is valid,

$$\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}=0 \qquad (2)$$

No heat can flow across the radial lines, which are lines of thermal symmetry, so the heat that is generated in the coil will leave the sector through the circumferential lines. Mathematically, the boundary conditions are

$$\frac{\partial T}{\partial n}=0 \text{ at the straight boundaries } (n=\text{coordinate direction normal to the boundary}) \qquad (3)$$

$$-K\left.\frac{\partial T}{\partial r}\right|_s=h(T_s-T_f) \text{ at the curved boundaries}$$

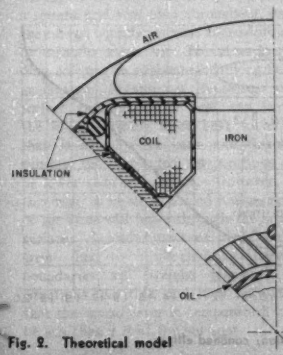


Fig. 2. Theoretical model

Solution

This is an extremely complicated problem to solve mathematically, for several reasons. The geometry of the coil is irregular; there are materials of different conductivities, the temperature at the boundary of the coil must simultaneously satisfy both Poisson's and Laplace's equations, and the boundary conditions at the curved external surfaces are difficult to handle. The only practical way to attack this problem is by a numerical method, and the "relaxation technique" is one possibility. This method is always laborious, however, and the added complications of this problem make any manual numerical procedure too time-consuming to be practical.

Instead the problem will be set up in the same manner as would be done in using the relaxation technique, but once the finite difference equations are derived, a digital computer will be used to solve them.

The finite difference equation to be solved at interior points of the coil is, for a square grid having rod length δ, (Fig. 3),

$$T_1+T_2+T_3+T_4-4T_0+\frac{Q\delta^2}{K_c}=0 \qquad (4)$$

the finite difference form of equation 1.

For an interior point of a nonheat-producing material of given conductivity, the finite difference form of Laplace's equation holds.

For all other points, such as boundary points of the coil and different materials, the finite difference equations are derived by applying the conservation of energy law to each point 0 and the four surrounding points, 1, 2, 3, 4 (Fig. 3). The general form of the finite difference equation at each point throughout the sector will be

$$aT_1+bT_2+cT_3+dT_4-eT_0+f=0 \qquad (5)$$

where the coefficients a, b, c, d, e, f are constants.

For a point within or on the boundary of the coil, f is nonzero; if one assumes that the heat generation rate Q is constant, then

$$f=\frac{Q\delta^2}{K_c} \qquad (6)$$

Actually, Q will be a linear function of temperature, since the resistance of the coil is a linear function. Then

$$\frac{Q\delta^2}{K_c}=AT+B, \quad A,B \text{ constants} \qquad (7)$$

and $f=B$.

In solving the problem by the manual relaxation method, first temperatures are estimated at each grid point, then the left side of equation 5 is evaluated for each point. The result will be generally a nonzero number which is called the residual at that point. The correct solution is obtained by altering temperatures according to the "relaxation pattern" until all residuals are reduced to a minimum. The relaxation pattern

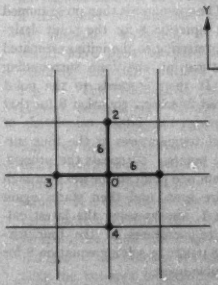


Fig. 3. Relaxation grid pattern

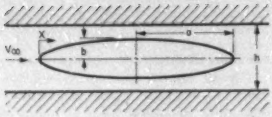


Fig. 4. Plane, potential flow past an ellipse in a straight channel

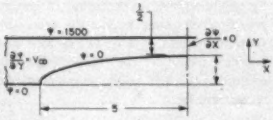


Fig. 5. Geometry for iteration solution of flow past ellipse

can be seen from equation 5. If one degree is added to the temperature at T_0, the residual at the point 0 will be reduced by e units; the residual at point 1 will be increased by c units, because as far as point 1 is concerned, the present point 0 looks like point 3; and so on. In performing the relaxation manually, the entire area is covered to eliminate the largest residuals first, and overcorrection is used to hasten the solution.

A digital computer can calculate residuals, but cannot easily be instructed to perform the relaxation as a human does it. The most efficient way for a computer to solve the problem is to perform iterations on equation 5. First equation 5 is written as

$$T_0 = \frac{aT_1 + bT_2 + cT_3 + dT_4 + f}{e} \qquad (8)$$

With each point in the region under study, there corresponds an identifying number, a set of coefficients a, b, c, d, e, f, and an initial estimated temperature. The computer is then programmed to solve equation 8 for the point designated 1, referring to the initial estimated temperatures at the four surrounding points. It then proceeds to the point designated 2, solving equation 8 for that point using the latest calculated or estimated temperatures at the four surrounding points. It repeats the process, going from one point to the next through the entire area, and then starts again at point 1, always using the latest calculated temperatures at the four surrounding points in solving equation 8 for a particular point.

A program has been written for the IBM *650* digital computer which will accommodate 704 grid points and 50 finite difference equations of the form of equation 8. The computer solves this equation at each point in 0.4 second, and performs calculations using ten digits. Advantages of speed and accuracy over the manual relaxation method are obvious.

Convergence Criteria

Some criterion must be established for halting the machine when it has reached an acceptable solution. As in manual relaxation, the magnitude of the largest residual may be used as this criterion, but this does not have much meaning when using coefficients in equations 5 or 8 which differ by more than an order of magnitude. A more practical criterion to use in machine computation is the change in values between two successive iterations. In the present program the machine stores the values at all points until the next succeeding iteration is finished. It then compares values at individual points, and if the sum of the magnitudes of differences for all points is less than a preassigned number for two successive iterations, the machine prints the last iteration and then halts. The preassigned number used as a criterion can, in the present program, be any power of 10, i.e., 10, 100, 1.0, 0.1, etc.

Rate of Convergence

The rate of convergence, or the speed with which the machine will furnish an acceptable answer for a given number of grid points and a given number of equations, will depend on the particular problem, of course. As a general rule, however, several factors play an important role in the convergence. One obvious factor is the initial estimated values at all nodal points: the closer these are to the correct solution the shorter will be the solution time. Other factors are the ordering of points, the proximity of the boundary to interior regions, and the boundary conditions. For example, points should be ordered such that regions that the machine studies first are those regions (1) within which the initial estimates are probably the poorest, hence changes will be the greatest, and (2) which affect other regions, such as boundaries. If the functional values at the boundaries are known, convergence will generally be faster than if it is necessary to use a relationship between the normal derivative at the boundary and an exterior value.

Verification of Program

The program was checked by solving a problem for which independently obtained solutions were available, and which at the same time involved most of the complications of the heat transfer problem within the rotor of an electric machine.

The problem is that of determining the pressure at points on the boundary of an air foil in plane potential flow without circulation in a straight channel, such as a wind tunnel, with a uniform velocity V_∞ at infinity upstream and downstream. The author has derived an integral equation for the velocity on the contour, and for purposes of checking this program the equation was solved for an air foil in the form of an ellipse, whose major axis is five times its minor axis, in a straight channel whose width is $1^1/_2$ times the minor axis. The major axis of the ellipse lies on the centerline of the channel, Fig. 4. Once the velocity is determined, Bernoulli's equation can be used to obtain the pressure.

In solving the same problem by the iteration method, the appropriate form

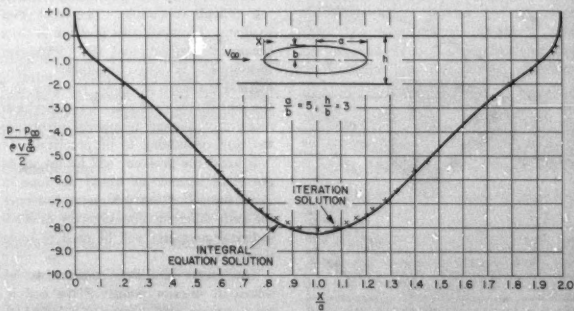


Fig. 6. Pressure distribution, confined ellipse

of the Laplace difference equation is written for the stream function at each grid point, the proper boundary conditions are inserted, and an initial estimated value for the stream function is assigned at each grid point. Because of the shape of the contour, symmetry exists about the two axes of the ellipse; therefore attention can be confined to one quadrant of Fig. 4.

The geometry used for solving the problem is shown in Fig. 5 with the boundary conditions. For grid points adjacent to the curved boundary, the proper coefficients for the difference equation 5 were obtained directly from Shaw[1] using his irregular star method for curved boundaries. A total of 45 equations and 403 grid points were used.

The grid points were ordered so that the computer solved the points adjacent to the contour before proceeding deeper into the fluid for each iteration. The computer stopped when two successive iterations showed a total change for all grid points of 10. The computer ran approximately 2½ hours before stopping.

The solutions obtained by the two independent methods are plotted in Fig. 6 as dimensionless pressure against the distance from the forward stagnation point. Agreement between the two solutions is excellent.

Temperatures in Rotor

In the rotor problem, temperatures were not known on the boundaries because the proportion of heat leaving through each surface was not known. Thus boundary conditions 3 were used. Also, the heat generation rate in the coil was taken to be a linear function of temperature, so equation 5 with condition 7 was solved at all grid points in the coil. Seven different conduction materials are present in the rotor. To simplify the writing of the finite difference equations, a square grid was used throughout, and the curved boundaries were approximated by straight segments. This does introduce an error in problems involving heat generation, because it increases the surface through which heat can leave the solid, allowing more heat to leave than is actually the case, and thereby causing the calculated temperatures to be lower than in practice. However, in this case, as in all oil-cooled rotors, most of the heat will leave through the small surface. A study[3] has been made of the error due to approximating curved boundaries by straight segments in situations like this, and it is estimated that the actual error in temperature will be less than 2 F at the hot spot.

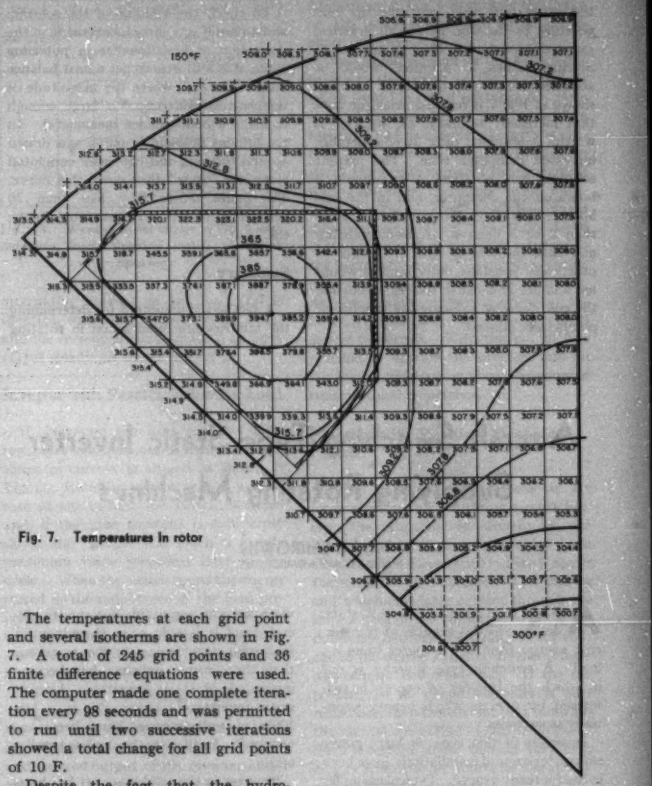


Fig. 7. Temperatures in rotor

The temperatures at each grid point and several isotherms are shown in Fig. 7. A total of 245 grid points and 36 finite difference equations were used. The computer made one complete iteration every 98 seconds and was permitted to run until two successive iterations showed a total change for all grid points of 10 F.

Despite the fact that the hydrodynamics problem discussed previously indicated that the program is good, the accuracy of the rotor solution should be checked if at all possible. In this problem the conservation of energy states that the heat generated in the coil should equal the heat leaving through both surfaces. However, due to the large heat transfer coefficient at the oil surface, the temperature difference at this surface will be small and an error in temperature of 1 F would cause a large error in the heat balance. In fact the high conductivity of the iron causes the temperature throughout the laminations to be practically constant, so an accurate heat balance will be difficult to obtain in electric machinery unless the computer is allowed to run until two successive iterations show a total change in temperature for all grid points of, say, 1 F. This is not considered practical for several reasons.

First, inaccuracies in the thermal conductivity values introduce some error in temperature, which will probably be larger than the error incurred by not having the machine run longer. Second, if this same problem had been solved by manual relaxation, a "solution" would have been accepted long before reaching the temperatures shown in Fig. 7. The residuals at points in the present solution are of the order of ±0.1, and acceptable residuals for the Laplace finite difference equation are ±2.0 when using manual relaxation. Finally, the objective must be kept in mind. It is desired to know the temperature pattern in the rotor, the hot-spot temperature, and the location of this hot spot. The solution of Fig. 7 indicates that the hot-spot temperature is around 400 F, that the temperature gradient in

the coil is quite steep, and that the temperature in the iron is, for all practical purposes, constant. Thus in most engineering studies, a "solution" would be accepted before the degree of accuracy shown in Fig. 7 is reached, with the note that increased accuracy, for what it is worth, is obtained only at considerably increased computer time. The present solution was obtained after almost 4 hours on the computer, and in the last hour temperatures did not change more than 2 or 3 degrees. Indeed, realizing now that the temperature in the iron is practically constant, later studies would use a larger grid size in the region outside the coil, thereby reducing the number of points to be studied.

To check the validity of the solution and to see if any errors were made in the derivation of equations or in punching cards, it is still wise to get a heat balance around the coil, where the magnitude of temperature differences is large enough to make a heat balance meaningful. In the present problem, a curve was drawn around the coil, and the heat computed along the "rods" that cross this curve. This figure differed by less than 20 per cent from the heat generated in the coil.

Summary

A method is presented for determining the temperature distribution in rotating electric equipment using an iteration technique that has been programmed for the IBM 650 digital computer. The program will accommodate 704 grid points and 50 finite difference equations. Using this program, temperatures in the rotor of an oil-cooled alternator were determined in less than 4 hours. Doing this same problem by manual relaxation would be economically unfeasible.

References

1. An Introduction to Relaxation Methods (book). F. S. Shaw. Dover Publications, Inc., New York, N. Y., 1953.

2. The Temperature Distribution in a Nuclear Reactor Fireced by Circular Cooling Channels. P. B. Richards. *Symposium Series*, American Institute of Chemical Engineers, New York, N. Y., vol. 50, no. 11, 1954.

Aircraft Switching-Type Static Inverter Supplying Rotating Machines

K. M. CHIRGWIN
ASSOCIATE MEMBER AIEE

A STATIC INVERTER is a static device which is capable of transferring energy from a d-c source to an a-c load. A switching-type inverter is one in which the transfer of energy is performed by devices which behave essentially as switches.

Inverters of this type, in large power ratings for ground installations, have been in use for many years.[1] Typical examples include the following. In electric traction systems, where it is sometimes necessary to return power from the d-c system to the a-c supply, this is often accomplished by means of a static inverter.[2] A combination static rectifier and inverter has often been used to transfer power between a-c power systems having different frequencies.[3] The equipment at the receiving end of a high-voltage d-c transmission line includes a static inverter.[4]

These inverters almost invariably use devices of the mercury-arc type, either an ignitron-type device or a grid-controlled mercury-pool rectifier. Mercury-arc devices are not acceptable for use on aircraft, and it is only since the advent of such solid-state devices as the transistor that the aircraft industry has become interested in static inverters. Aircraft static inverters of a few 10's to a few 100's of volt-amperes have been announced. In all cases they utilize transistors.

Progress in the solid-state art has been very rapid in the past few years and solid-state devices of very much greater power-handling capability will be available in the near future. It is probable, therefore, that in the next few years aircraft-type static inverters of much higher power ratings will become available.

The purpose of this paper is to examine the probable behavior of these future higher powered static inverters with aircraft-type loads, particularly when this load includes rotating machines, and to examine some of the problems that may be anticipated.

Sine-Wave versus Square-Wave Inverters

The requirement for a sine-wave output from the static inverter arises because present aircraft a-c systems, whether supplied from an alternator or from a rotating inverter, have a sine-wave power supply, and consequently present aircraft utilization equipment is designed for sine-wave input.

Sine-wave inverters have been announced in small ratings in which either a pair of transistors is operated as a class *B* amplifier, or a square-wave transistor inverter is used with some form of waveshaping or filtering. Because a large part of the difference between the square wave of the switched d-c source and the sine-wave output of such an inverter is absorbed in the waveshaping components or in the solid-state device itself, these inverters are inefficient and produce considerable losses in the inverter components, which have to be dissipated.

The natural shape of wave for a switching-type static inverter is some form of square or stepped wave. Waves of this type are not generally applicable to aircraft loads as they are presently known. However, there are certain loads which could use a square wave to advantage. Such, for example, is a high-voltage d-c power supply, where the square wave reduces the amount of smoothing needed on the d-c output.[6]

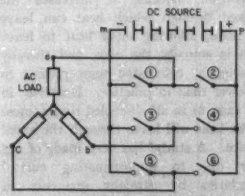


Fig. 1. Switching inverter 3-phase bridge circuit

Paper 58-828, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 17, 1958; made available for printing April 10, 1958.

K. M. Chirgwin is with Jack & Heintz, Inc., Cleveland, Ohio.

Switching-Type Inverter

As will be seen later, it is possible to operate a switching-type inverter into a rotating machine load and provide essentially sine-wave voltage, without special output filters, and at the same time keep the internal losses in the inverter quite small. For this reason the term "switching-type inverter" is used rather than "square-wave inverter."

Because a switching-type inverter has high efficiency and a minimum dissipation requirement in the switching device, it is reasonable to assume that the future higher power inverters will be of the switching type. All subsequent discussion is confined to the switching-type inverter.

An approximation of the behavior of this inverter using solid-state switching devices can be had by assuming the switching device to be a perfect switch, that is, a device which has zero impedance when "on," infinite impedance when "off," and zero opening and closing times. It is on the basis of such a perfect switch that an examination is made of the inverter behavior.

Inverter with Passive Resistive Load

A switching inverter using a full-wave bridge circuit is illustrated in Fig. 1. It consists of a d-c source shown as a battery, six switches which are opened and closed in a definite sequence, and an a-c load. The wave of voltage output of this full-wave circuit is illustrated in Fig. 2 when the switches are each closed for 120 electrical degrees.

If the a-c load on this inverter is pure resistance, then the waveshape of the

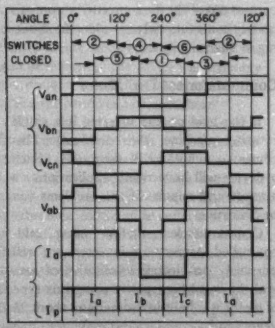


Fig. 2. Waveforms for circuit of Fig. 1 with resistive load

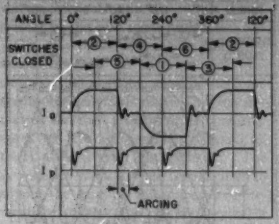


Fig. 3. Waveforms for circuit of Fig. 1 with inductive load, showing arcing at switch opening

alternating current in the load (I_a) is identical with that of the voltage (V_{an}) and the current drawn from the d-c source (I_p) is steady, as shown in Fig. 2.

Inverter with Passive Inductive Load

If, however, the load contains both resistance and inductance, then the waveshape of current is altered as shown in Fig. 3. Here the current rises at a finite rate at the instant the switch is closed and, if the time constant is sufficiently short, the current will reach a steady maximum value sometime later in the cycle. When the switch opens the energy stored in the inductance of the load prevents the current from dropping instantaneously to zero and the inductively stored energy is dissipated in an arc which is drawn at the switch.

Operation in this manner in an actual inverter, may cause immediate failure of the solid-state device, or may cause heating of the device with loss of efficiency and reduced output of the inverter and a spike on the voltage waveform. The oscillogram of Fig. 4 illustrates this spike on the voltage waveform of a single-phase inverter.

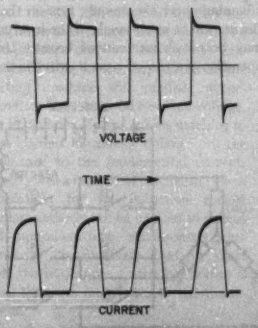


Fig. 4. Oscillogram showing current and voltage waveforms of a single-phase switching inverter with inductive load

One possible remedy for this situation is shown in the circuit of Fig. 5, which is taken from Fig. 18 of reference 6. Here each of the switches is shunted with a rectifier which is capable of carrying current in a direction opposite to that in which the switch normally carries current during inverter operation. The waveforms in Fig. 6 show how this circuit modification operates. The energy stored in the inductance of the load is returned to the battery by way of the rectifier, during a short period following switch opening. At the instant of switch opening, the battery current drops to zero and returns slowly to normal. With this circuit it is possible to operate the inverter into an a-c load which has any combination of resistance and inductance.

Inverter with Rotating Machine Load

When a static inverter feeds an a-c load which contains rotating a-c machinery, the situation is altered by the presence of a 3-phase sinusoidal back electromotive force (emf). Fig. 7 shows the situation in this case. The driving emf which circulates current through the impedances of the circuit elements is the difference between the steady direct voltage of the source V_{pn} and the instantaneous counter emf wave E_{pn}. If the impedance of all circuit elements is pure resistance, then the current that flows has the same shape as that of this driving emf. Such a current would transfer very little energy from the d-c source to the fundamental of the a-c system.

In order to get good transfer of energy to the fundamental of the a-c system, it is necessary to insert inductance on the d-c side of the switches, which will absorb the differences between the steady direct

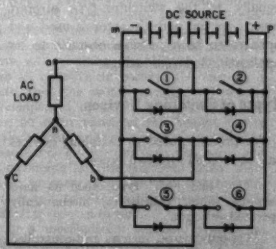


Fig. 5. Switching inverter 3-phase bridge circuit with shunting rectifiers

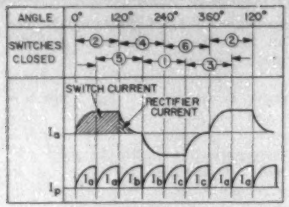


Fig. 6. Waveforms for circuit of Fig. 5 with inductive load

voltage of the source and the instantaneous counter emf wave, and will constrain the current flowing out of the d-c source to remain approximately constant at all points in the cycle. The circuit is shown in Fig. 8 and the waveforms in Fig. 9.

Comparison with a Mercury-Arc Inverter

The circuit arrangement and method of operation now arrived at are very similar to those of a grid-controlled mercury-arc inverter operating into an a-c system. There is a considerable body of knowledge existing in this field[3,7] and much of it can be applied directly to the switching-type aircraft inverter.

One very important characteristic of the mercury-arc inverter is that it cannot supply reactive volt-amperes to its a-c load, and in fact must have considerable reactive volt-amperes supplied to it from the load. The amount of the reactive volt-ampere requirement of the mercury-arc inverter ranges from approximately 50% to more than 100% of its watts output depending upon the frequency of the output, the deionization time of the inverter tubes, the inductance of the a-c output transformer windings (or its equivalent), the overload capacity for which the inverter must be designed, and other factors. This limitation applies equally to the switching-type aircraft inverter when a d-c reactor is used and appears to be a possible obstacle to its application in larger ratings.

Types of Switching Devices

For present purposes switching devices can be divided into two types:

1. "On" and "off" type, such as hard vacuum tubes, transistors, mechanically operated switches.

2. "On-only" type, such as grid-controlled mercury-arc rectifiers, ignitrons, thyratrons, grid-controlled gas-filled tubes, solid-state controlled rectifiers.

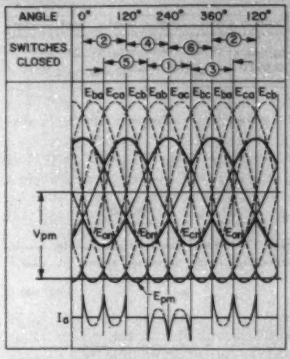


Fig. 7. Waveforms for circuit of Fig. 1 with sinusoidal back emf and resistive circuit elements

The basic difference between these two types of switches is that type 1 devices can be switched on or off at will by means of the control element; whereas type 2 can be switched on by means of the control element, but will only cease conduction when the power current is forced to zero by other elements in the power circuit.

The switching device in the grid-controlled mercury-arc inverter is, of course, of the on-only type, while that of the switching-type aircraft inverter may be of either type.

The Appendix shows that unless an on-off type of switching device is to be operated in the "arcing" mode (internal power dissipation), inverter operation is essentially the same for either type of switch.

This Appendix shows, however, that the on-off type of switch may have some advantage over the on-only type in that the margin or safety angle of the inverter may be somewhat reduced because the operation of an on-off type of switch in the

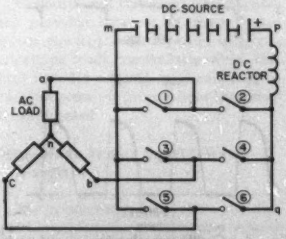


Fig. 8. Switching inverter 3-phase bridge circuit with d-c reactor

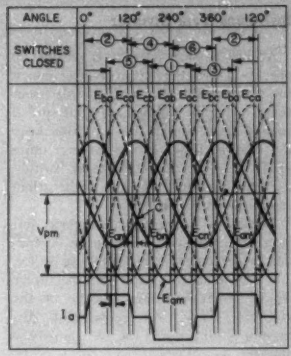


Fig. 9. Waveforms for circuit of Fig. 8 with sinusoidal back emf, showing overlap of switches

arcing mode, which occurs when the safety angle is exceeded, is much less disastrous than the arc-through which occurs when an inverter using on-only type switches is operated beyond its safety angle. If arcing mode operation is excluded, then it is still true, with either type of switch, that a switching-type static inverter with d-c reactor must be supplied with reactive volt-amperes by its a-c load. However, if the safety or margin angle can be reduced, then the amount of this reactive volt-ampere requirement is also reduced.

Use of rectifiers shunting the solid-state switches (Fig. 5) was found to be an effective way of supplying reactive volt-amperes to a passive load. This circuit arrangement is not applicable to rotating machine loads because the d-c reactor, required to give good transfer of power to the rotating machine, constrains the current out of the d-c source to be approximately constant, and so renders the shunting rectifiers inoperative.

Commutatorless D-C Motor

If the load on the inverter is a single rotating machine, the combination becomes a commutatorless d-c motor. Such a device will have many applications on aircraft and missiles if the problems can be overcome.

Commutatorless motors using grid-controlled mercury-arc inverters are well known.[8] An electric locomotive of several thousand horsepower using this type of motor was built in Switzerland about 25 years ago.[9]

References 6, 10, and 11 deal with commutatorless d-c motors using transistors

as the inverting device. The circuit used in references 10 and 11 is a 3-phase half-wave circuit. This configuration is undesirable since it causes direct currents to flow in the a-c windings of the motor, which cause the motor to act as an eddy-current brake. Reference 6 uses a 3-phase bridge circuit which overcomes this objection.

All three of these references use induction motors as the rotating machine so the inverter must supply reactive volt-ampere to the motor. As has already been seen, this requires the transistors to operate in the arcing mode which gives rise to appreciable internal losses in the transistors and is therefore undesirable, but may be tolerable for small units.

A solution to this problem is to use a synchronous machine as the motor. (It is significant that the Swiss locomotive referred to used synchronous motors.) When running, the excitation of this synchronous motor can be adjusted to supply leading volt-amperes to the inverter to enable it to operate with an adequate margin angle at all times so that arcing mode operation need never occur during running.

Motor Starting Problems

When starting a commutatorless d-c motor it is possible to reduce the frequency of the inverter to some low value, then bring the motor up to the low synchronous speed as an induction motor, and synchronize it. Once synchronized, the motor can supply the reactive volt-ampere requirements of the inverter. During the period the motor is in operation as an induction motor, the motor requires a considerable amount of reactive volt-amperes, which will require either the inclusion of large static capacitors in the circuit, or arcing mode operation of the solid-state device during the starting period. However, if arcing mode operation is used the heat developed in the device will probably not present too serious a problem, providing the period can be kept short.

When the inverter must operate at a fixed frequency because of the requirements of other items in its load, the starting of a synchronous or an induction motor poses serious problems. Starting a synchronous motor as an induction motor or starting an induction motor from a fixed frequency supply requires large amounts of reactive volt-amperes during the starting period. Arcing mode operation of the solid-state device during this period presents a serious problem because of the large amount of losses developed in it. Fixed capacitors of the required volt-ampere rating in the output circuit present weight and space problems.

One way of avoiding these difficulties would be with an inverter and synchronous motor combination, in which the synchronous motor is started at a low frequency and then brought up to operating speed and frequency and kept running all the time the inverter is in operation. The synchronous motor would then be used to supply the reactive volt-ampere requirement for starting and running any other motors which form part of the inverter load.

A-C Waveform

The waveform of the output voltage of a switching inverter operated on resistive or resistive inductive loads without a d-c reactor is essentially a square or stepped wave. There are a number of ways of making this wave more nearly sinusoidal which are outside the scope of this paper.

The switching-type inverter, having inductance between the d-c source and the switches, and supplying rotating machines would give a pure sine-wave output voltage, provided the back emf of the machine is a pure sine wave and provided the impedance voltage drops do not distort the terminal voltage. The back emf of the machine can be made a very close approximation to a sine wave, but the impedance drops do lead to some departure from the pure sine wave.

The waveforms of Fig. 9 show that the current waveform in such a system is approximately a trapezoid with about 120 electrical degrees width. Such a wave contains all odd current harmonics. Triplen harmonics are small in magnitude when the width is near 120 degreees and the amplitude of the other harmonics will in general, decrease with increase of order, so that the 5th, 7th, 11th, and 13th are the most important current harmonics. The voltage drops due to these current harmonics flowing in the impedance of the rotating machine will produce departures from the sinusoidal terminal voltage. The impedance of the rotating machine to these current harmonics differs from the impedance to the fundamental current. Alger[12] gives a method of calculating these impedances for an induction motor. Similar methods can be applied to salient and nonsalient-pole synchronous machines. It is fortunate that in most cases the impedance to the harmonics is lower than the impedance to the fundamental.

Rough calculations using this method indicate that *MIL-G-6099A* requirements of harmonic content of the terminal voltage can probably be met by a switching inverter and synchronous motor combination, but there will probably be some weight penalty.

Conclusions

Advances in the solid-state art hold promise of components which will make high-power aircraft-type static inverters possible in the near future. It is presumed that these inverters will be of the square-wave or switching type. By representing the expected solid-state device by a perfect switch it is possible to make some predictions about these inverters and to apply much of the knowledge presently available from the field of grid-controlled mercury-arc inverters.

In particular, it is concluded that the switching inverter supplying rotating machines should have a d-c reactor and a leading power factor a-c load (a synchronous motor, for example).

The solid-state switching device can be either of the "on-off" type (such as a transistor), or of the "on-only" type (such as solid-state controlled rectifier).

A switching inverter feeding a rotating machine will have approximately sinusoidal voltage output waveform and approximately trapezoidal current output waveform.

The motor starting capability of a switching inverter is rather poor. If the motor is a synchronous machine and the inverter output frequency can be reduced to a low value, the starting problem is not too severe. If the inverter output frequency must be held constant, motor starting becomes difficult. A possible solution is to have a synchronous machine, which runs all the time, across the inverter output.

Appendix

Switch Operating Characteristics

When a switch is open the current and the inductively stored energy in its associated circuit must both be zero. If at the instant before switch opening the stored energy is zero, either because the circuit components are all resistive or because the current is zero, then the switch opening will not be accompanied by any energy transfer. If, on the other hand, at the instant before switch opening there is energy stored in the associated circuit (current not zero and components not all resistive), then there must be energy transfer, and switch opening is not complete until the energy in the circuit is reduced to zero.

In a mechanically operated switch the circuit energy is reduced to zero by being dissipated in the arc which is drawn between the separating contacts. In a transistor or vacuum tube operated as a switch

circuit energy is dissipated inside the device. It is a requirement for an on-off type of switch that it be capable of dissipating energy.

Operation of an on-off type of switch which involves dissipation of stored energy within the switch itself is referred to as "arcing mode" operation.

An on-only type of switch does not have this energy dissipating capability and so arcing mode operation is impossible. Opening the circuit with an on-only type of switch requires that the current in the circuit be reduced to zero (i.e., the energy stored must be zero) before the switch will open.

It is possible, and often desirable, to open an on-off type of switch when the circuit energy is zero. A mechanical rectifier is a good example of this case.

Inverter Operation with an A-C Back EMF

Inverters for these conditions have been built and operated successfully using grid-controlled mercury-arc devices as the switch. These are, of course, on-only type switching devices.

An important requirement for successful operation of an inverter which uses an on-only type of switching device is that the switch in one circuit be opened while the instantaneous value of the alternating back emf of its circuit is greater in magnitude than that of the next circuit in the switching sequence. That is, the switch must be open before reaching point *C* indicated in Fig. 9. If this requirement is not met, the circuit will remain closed at point *C* and an arc-through will result, which means that the circuit will continue to conduct for at least 3/4 of a cycle. During part of this time, the voltage of the d-c source and the back emf of the a-c machine will be aiding each other in driving a current through the impedances of the circuit elements. This, of course, represents a short circuit on both the a-c and d-c sources, and can be disastrous.

In order to avoid any possibility of an arc-through it is necessary to have a margin or safety angle between the point where the switch is open and point *C*. The need for this margin arises mainly from two causes; (1) the mercury arc takes a finite time to deionize and become nonconducting after the current has reached zero; and (2) the angle of overlap between adjacent circuits in the switching sequence increases with the load current and some allowance must be made for overloads.

When an inverter with an alternating back emf is built with on-off type of switching device, control is not lost if the circuit is not opened at point *C* since the switch, if adequately designed, can still open the circuit and permit the normal switching sequence to proceed. This is undesirable as the normal method of operation since it involves arcing mode operation of the switch with its attendant reduction of the efficiency of the inverter and increased cooling requirements of the switch.

It is reasonable, however, to reduce the margin angle in an inverter using on-off type switching devices, since loss of margin due to overload or other emergency conditions does not lead to a disastrous arc-through, but merely to increased heating of the switch, which can be tolerated if it is not too severe and if it persists for short periods only.

The advantage to be gained by reducing the margin angle is that the circuit is now closed and opened later in the cycle of the alternating emf, and the current is now more nearly 180 degrees out of phase with the emf and the leading volt-ampere requirement of the inverter is reduced.

References

1. Mercury-Arc Power Converters in North America, AIEE Committee Report. *AIEE Transactions* vol. 67, pt. II, 1948, pp. 1031–59.

2. High Voltage Ignitron Rectifiers and Inverters for Railroad Service, J. L. Boyer, C. G. Hagensick. *Ibid.*, vol. 65, July 1946, pp. 463–70.

3. The Electonic Converter for Exchange of Power, F. W. Cramer, L. W. Morton, A. G. Darling. *Ibid.*, vol. 63, 1944, pp. 1059–69.

4. The Theory of the Control Problem of H. V. DC Transmission with Rectifiers and Inverters in Bridge Circuit, F. Busemann. *Report Z/T74*, Electrical and Allied Industries Research Association, London, England.

5. A New Self-Excited Square-Wave Transistor Power Oscillator, G. C. Urchin, W. O. Taylor. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, Jan. 1955, p. 99.

6. Direct Engine-Driven Unconventional Electric System, K. Martinez. National Aeronautics Meeting, Society of Automotive Engineers, New York, N. Y., 1957.

7. Analysis of a 3-Phase Inverter with Resistive Load, Robert E. Turkington. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 1055–61.

8. The "Ignitron" Type of Inverter, C. F. Wagner, L. R. Ludwig. *Ibid.*, vol. 53, Oct. 1934, pp. 1384–88.

9. Passenger Locomotive Converted to Rectifier Control, E. Kern. *Elektrische Bahnen*, Berlin, Germany, vol. 7, 1931, p. 313.

10. Alternating Current Motor, Carl-Erik Granquist. *U. S. Patent no. 2,810,843*, Oct. 22, 1957.

11. Transistor Circuit Engineering (book), R. F. Shea. John Wiley & Sons, Inc., New York, N. Y., 1957, pp. 428.

12. Nature of the Polyphase Induction Machine (book), P. L. Alger. *Ibid.*, 1956.

Design of an Aircraft Power-Type Circuit Breaker for 600 F Ambient

B. O. AUSTIN
FELLOW AIEE

PROJECTS in the 600 F (degrees Fahrenheit) ambient temperature area have recently assumed major importance in many new aircraft development programs. Significant progress had been made in the design and prototyping of a high-temperature aircraft power-type circuit breaker. The interplay of many ideas in the mechanical, electrical, and materials fields contributed to the effort to solve the difficult and strange design problems which were encountered in this project.

Application of the new breaker follows, to some degree, the pattern of low-temperature breakers now in use. Experience with them has guided this program to some extent.

The breaker is of the latch type, rated at 225 amperes, 480 volts, 3 phase, 400 cycles, 1,000 hours' life, with target performance of 4,000 amperes interrupting capacity, 5,000 interruptions at rated current, closing time 0.05 second or less, trip time 0.01 second or less, 75,000 feet altitude, and 600 F ambient temperature. Operation is controlled by signals from other sources with no means for overload sensing.

In addition to the new temperature requirements, other environmental requirements were added. These were corona suppression, ozone, acoustic noise, and temperature shock. Corona suppression requires creepage distances between live parts in the order of $1^1/_4$ inches. Tests at altitude have demonstrated this. Operation is to be satisfactory in an ozone concentration of 0.015 per cent in atmosphere by volume. The acoustic noise specification requires operation at a level of 150 decibels, through a wide frequency range. Thermal shock resistance is required when temperature changes at a rate of 160 F per minute. Environment conditions, such as altitude, shock acceleration, vibration, salt, spray, humidity, fungus, explosion, sand, and dust, were substantially the same as those for breakers in present use.

Paper 58-832, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 10, 1958; made available for printing April 10, 1958.

B. O. Austin is with the Westinghouse Electric Corporation, Lima, Ohio.

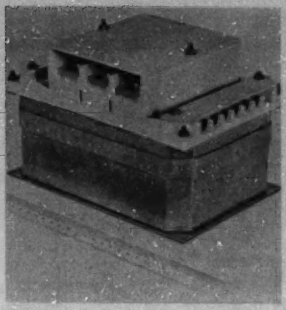


Fig. 1. Aircraft power-type circuit breaker, 600 F ambient

Design Procedures and Description

Many problems requiring new concepts and structural arrangements were recognized in the beginning. However, many others were uncovered during the complete analysis.

Several basic conditions were set up by the author for the purpose of producing a satisfactory product in a reasonable length of time; Figs. 1, 2, and 3. Perhaps the most important condition was that known materials, new or old, must be used. Also, the use of the following were eliminated.

1. Bearings.
2. Pivot pins.
3. Encased plungers.
4. Materials with excessive percentage of relaxation at temperatures below 600 F.
5. Brazed joints in the electric circuits.
6. Load currents through studs.
7. Delicate adjustments.

All of these objectives were realized in the final design of the product.

The first step in the design was to select the means of interrupting the power circuit. Voltages to be handled were double the voltages of the low-temperature design now in use. The logical solution was to double the number of gaps to be opened. This provided four breaks in each phase. See Fig. 4.

The next consideration was the mounting of the contact mechanisms. Here again, previous successful designs were followed, except that because of high temperature, ceramic pockets rather than molded ones were provided in the breaker base which enclosed the contact parts.

Materials for making the contacts was another selection to be made. In the final analysis, and after consultations with engineers from the contact manufacturers, silver-cadmium oxide was selected. Subsequent tests at the required temperatures and load currents verified that the decision was correct.

Brazing of contact buttons on other materials has not been completely satisfactory at normal temperatures. They have been known to fall off or offer high resistance. A conclusion was reached to make these parts of solid silver-cadmium oxide. This material has low contact resistance and low antifreeze characteristics which make it attractive for this application.

The theories of arc interruption are much too involved for a complete discussion in this paper. However, briefly, the major agents for deionizing and extinguishing electric arcs in this breaker are

1. Multiple gaps.
2. Pressure formed by the hot arc in the ceramic pockets (making the breaker much less sensitive to altitude conditions).
3. Cooling of the arc by the ceramic pockets and contact parts.

Ceramic was selected for the main bases and contact pockets, push rods, caps, solenoid end washers, and wire insulation. Ability to meet all the requirements that are involved, low cost, and the desired molding characteristics of porcelain were the deciding factors.

The problem of operating the moving contact bar mechanism was solved by adopting the straight-line push-rod motion used in the low-temperature aircraft breakers, except that nickel-chromium-iron-titanium alloy is used throughout the breaker design for all springs. There are two helical springs in this assembly: a contact compression spring and the contact-opening spring. Steatite ceramic push rods and caps are used. See Fig. 5. The assembly is held together by a through-stud and locknut of nickel-chromium alloy. Each of the six contact assemblies moves independently.

Nickel-chromium-iron-titanium alloy was selected for making springs because of its ability to meet the ambient temperature, thermal shock, and ozone requirements, with a very low percentage of relaxation or damage. No other material was found to possess this characteristic to an equal degree.

Nickel-chromium was selected for the structural parts, the covers, and for standard and special fasteners. The material is very stable under the requirements, readily available, low in cost, and not too difficult to work.

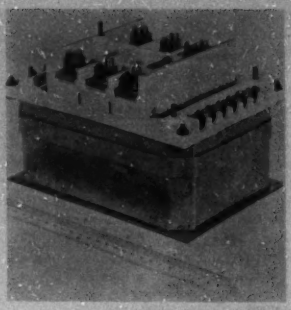


Fig. 2. Circuit breaker with main and control covers removed

Titanium and its alloys were seriously considered for these parts in order to save weight. The main reason titanium was not used was due to its relaxation at temperatures lower than permissible in this design. A usable alloy of this material may become available at a later date.

The next design solution to achieve was the mechanical means for operating the six main contact push rods and eight auxiliary contacts or interlocks. This required an interplay of many ideas to meet all the requirements. Vibration and shock were the outstanding problems involved in this phase of the design.

From all the concepts and structures considered, an armature having a straight-line motion seemed to be the simplest and most adaptable. All 14 components to be operated could be placed approximately in one plane. In addition, separate adjustments could be provided for each component to be operated. A flat-type armature with straight-line motion and individual adjustments was adopted. See Fig. 6.

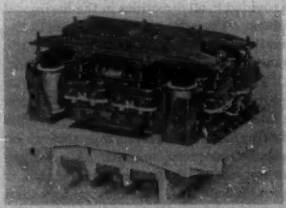


Fig. 3. Internal view of breaker showing solenoids, interlocks, ceramic bases, and armature

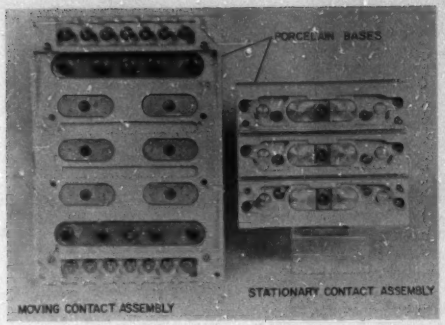


Fig. 4. Ceramic bases separated showing main contact assemblies

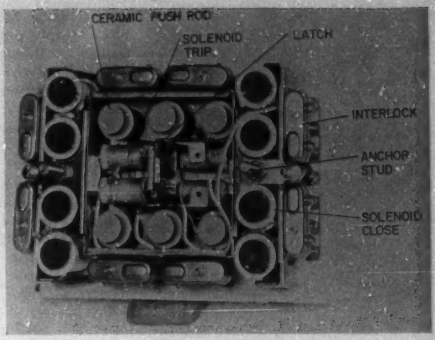


Fig. 5. View of breaker showing main armature removed

Closing magnetic forces were calculated to be about 38 pounds, initial forces about 14 pounds.

Means for producing the necessary magnetic forces was considered from many angles. The best mechanical and magnetic structure seemed to be the use of four small solenoid magnets at each end of the breaker. There are several advantages in this arrangement:

1. A symmetrical structure is achieved.
2. It is adaptable to straight-line motion.
3. Solenoid coils can be series-paralleled.
4. More efficient uses made of space.
5. It is easier to design small high-temperature solenoids.

With the type of armature and magnetic structure determined, a novel support for the armature was designed. A flat spring extends to each end of the breaker. The ends of this spring are rigidly attached to a stud and the armature is attached to the center of the spring. Two slight offsets in the spring permit the armature to move up and down freely. Stability against vibration and shock is provided in all directions in the horizontal plane. The armature is restrained in the vertical directions by the springs in the contact mechanisms.

The next design consideration was to provide a satisfactory latch and latch plate at the center of the armature. The latch is mounted on a flat spring which provides ample movement. A pivot pin is not used. Operation of the latch is provided by two solenoids mounted on each side of the latch. Mounted on the latch are two soft iron armatures. Material used for making the latch is nickel-chromium.

Latch plates, a part on which the latch shoulder rests, posed another problem at high temperatures and other environmental conditions. Several materials, including ceramics, were investigated.

Tungsten-carbide was selected because of its hardness, smooth finish, and its ability to withstand the environment without deterioration. It proved to be an excellent material for latch surfaces. Ceramics and chromium were tried but discarded because they produced more friction during trip operation.

Making a solenoid for the close and trip electromagnets which would withstand the temperature limits of −65 F to 600 F ambient was one of the more difficult undertakings. Magnet wires or straps with 600–800 F insulation were not available. However, solenoid coils were designed by using known materials. In this design, fine silver ribbon, 10 mils thick for close coils and 5 mils thick for trip coils, both 100 mils wide, are used for the solenoids. This material is wound on edge, one layer deep, to form the winding for the close and trip coils. The coils are then treated with a high-temperature silicone varnish. Before the varnish is completely dry, the coils are opened up and powdered between turns with magnesium oxide, and then placed on borosilicate glass tubes for insulation from the cores. The coils and glass tubes are then baked at a low temperature. The coils and glass tubes, and ceramic end washers, are next placed on the core, dipped in silicone varnish, and baked at a high temperature which burns out most of the silicone varnish and leaves the magnesium oxide between the turns for insulation.

Materials Requirements for Solenoids

Fine silver was selected for its ability to withstand the environmental conditions specified and for the compactness of the

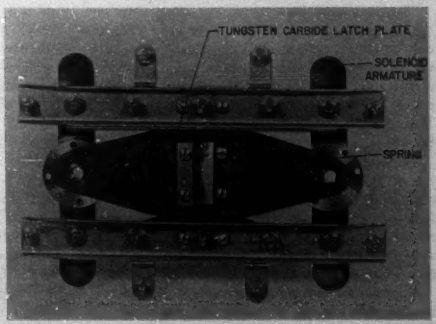


Fig. 6. Main armature showing flat spring support

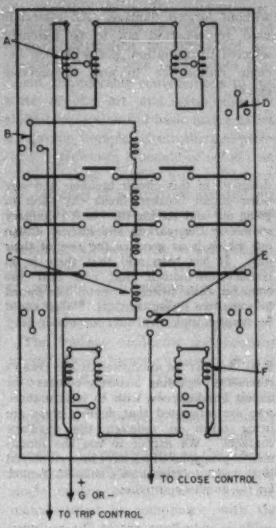


Fig. 7. Main and control circuit diagram

A—Close solenoids
B—Interlock closed by main armature in its closed position
C—Trip solenoids
D—Spare interlocks
E—Interlock closed by the latch in the unlatched position
F—Close solenoids

completed solenoids. Other materials, such as nickel-plated copper and aluminum, were considered. However, the nickel plating on the copper is not a very good conductor and it occupies approximately 2 mils between turns, which makes the nickel-plated copper coil larger than the fine silver coil. Also, copper itself will not stand the specified temperatures. It deteriorates by oxidation. The aluminum coil is also larger than the silver, thus requiring larger and heavier magnetic circuits as well as more space in the circuit breaker.

Borosilicate glass was selected for its ability to withstand the specified conditions. This material is readily available and can be manufactured to rather close tolerances in tubular form, which is desirable in the construction of the solenoids. The material is used as an insulator between the coil and its core.

Magnesium oxide was selected for use between the turns of the single-layer edge-wound fine-silver solenoids. This material was found to be adaptable to powdering of the coil after a varnish treatment and to withstand the temperatures even after the silicone varnish is burned out.

Silicone varnish was selected more as a means for holding the magnesium oxide between the turns of the solenoid until the solenoid is clamped into the space provided in the breaker.

Magnetic materials for the solenoids did not present a problem. Pure iron was used because of its good magnetic characteristics, plus a Curie point which is well above the operating ambient temperatures specified.

Terminals and Bases

The main terminal studs, of which there are six; the stationary contacts; and the bridge between the two bridge contact assemblies are fitted to one of the porcelain bases. Stationary terminal contacts are bolted directly to the main terminal stud bases, thus avoiding the carrying of current through the stud itself. All other items, including control terminals and covers and moving contact bridge assemblies, are mounted on the second base. Main terminals are made of copper-chromium alloy, the cover and structural parts of nickel-chromium, the terminal covers of glass-bonded mica, and the internal interconnecting wiring of fine silver with ceramic bead insulation. See Figs. 2 and 3.

Glass-bonded mica was selected for the terminal covers because the temperature never exceeds 600 F in this particular location. The glass-bonded mica used is a new version of this material and will stand temperatures somewhat higher than required.

Copper-chromium alloy was selected for its high conductivity and high strength for the main terminal studs. The control terminal studs are nickel-chromium.

Operating Characteristics

The breaker is flexible in its application. It can be used in 3-phase a-c 480-volt 400-cycle circuits. It also may be used in 250-volt d-c circuits. In the case of d-c circuits, all contacts are connected in series. In fact, the first application will be to control the power circuit of a 160-horsepower series motor.

Electrical interlocking of the circuits of this breaker is different from that used in present circuit breakers. This was done to improve the reliability of the breaker as well as to avoid some delicate adjustments. Previous breakers were electrically interlocked so that the cutout interlock contacts to the close coil and trip coil were operated by the main armature, requiring a fine adjustment of the close coil interlock contact.

A feature incorporated in this breaker is that the interlock in the circuit of the close coil is operated by the latch movement. The breaker must be latched before the close coil is cut off. On the other hand, the trip coil stays closed until the latch is opened and the movement of the main armature opens the interlock circuit to the trip coil. This is not a high-temperature requirement; it is considered an interlocking improvement.

The close coils and trip coils are energized momentarily to close or trip the breaker. See Fig. 7 for diagram and circuitry.

Testing and Testing Facilities

Subjecting the complete breaker to a test in a 600 F ambient temperature for a period of 2 hours revealed that none of the selected materials were damaged. Power current was not conducted by the breaker during this test. Magnetic operation was satisfactory at this ambient temperature.

Testing of contact material has proved that it is satisfactory for the temperatures and load currents specified. All contact parts and supports are of the same material. Brazing is not used.

D-c testing at 250 volts and 550 amperes indicates that the series contact arrangement is satisfactory for interrupting d-c power. This is also an indirect proof that the breaker will handle the 480 volts. A-c power circuits are much easier to interrupt than are d-c power circuits.

As testing facilities are very limited for making all the required tests, it will be some time before a complete test report can be submitted. See Fig. 7 for the circuit diagram.

Summary and Comments

This paper reports the design procedures and progress made in the development of a 600 F ambient aircraft power-type circuit breaker. The methods used in design solution of the mechanical, electrical, and material problems are reported. Special attention is called to the general structures, solenoids, armature support, contact materials, and ceramic bases.

It is to be noted that the increased temperatures, added environmentals, and changes in the existing environmental requirements are very decided factors in

the increasing weight. The weight of the first unit was 18.5 pounds, but further design refinement will reduce this to 17 pounds or less.

Design procedures clearly demonstrate that with creative thinking many high-temperature devices can be designed with available materials.

The circuit breaker is expected to withstand high-intensity nuclear radiation without serious damage. All materials used are reported not to be seriously affected by radiation, nor will they become radioactive to any appreciable extent.

Discussion

T. B. Owen (Douglas Aircraft Company, Santa Monica, Calif.): It is very encouraging to know that the "thermal thicket" is as penetrable on the electrical design level as it is on the aero-dynamic design level, and at not too great a weight penalty for the performance obtained. However, one point I would like to deal with in regard to this design is that it is incomplete.

By this I mean that although Mr. Austin has done a tremendous job on the contacts, operating coil, and all of the other parts of the circuit breaker concerned with the high-current portions, he has gone out and bought, not designed, the auxiliary switches. I know of very few circuit-breaker applications where the auxiliary switches are not at least as important as the main contacts; (we all use them to short-circuit current transformers, etc.) and it is essential that this short-circuiting be properly done. If it is not, then we will have system trouble, low or high voltages, generators that will not stay paralleled, and all the rest. I believe that since the auxiliary contacts are at least as important as the main contacts, the same design consideration should be given them.

Until our circuit-breaker designers design auxiliary contacts that are something other than a pair of 1/8-inch-diameter silver contacts rubbing together, we are going to experience difficulties regardless of the inert atmosphere in which they operate.

I wish to cast no aspersions on the auxiliary switch used as it is undoubtedly a good one, but no snap switch is good enough for the integrity of a power system to depend on it. I would like to see the designers of this circuit breaker, and any other circuit breaker, break the habit of going out and buying the "best" auxiliary switch on the market, and instead design one which is as good as the rest of their unit. It has been said that there is no contact mechanism that can beat the common knife switch, a metal bar forced between two spring contacts. What could be simpler, and why can't we obtain it?

B. O. Austin: In an answer to Mr. Owen's comments regarding auxiliary contacts for circuit breakers, we wish to inform those who are interested that definite steps are being taken to redesign the auxiliary contacts. We expect to use the fundamental principles which have been employed in the past to guarantee a reliable contact for the auxiliary switches.

Conductors and Conductor Protection

R. STEINER
ASSOCIATE MEMBER AIEE

IT is a purpose of this paper to note the principles underlying the establishment of current-carrying capacities of electric conductors including those for air-borne applications. Another purpose is to determine the concepts regulating the requirements for the protection of the conductors. Further objectives of this presentation comprise the consolidation of conductor performance and conductor protection requirements, the distinction between laboratory and field significance of these data, and the evaluation of the respective product characteristics in comparison with specification requirements.

The rapid development of the electrical art rarely allowed for the timely and concerted preparation of performance specifications for both electric conductors and the devices to protect them. This situation became more aggravated in times of national emergencies when the manpower necessary for the writing of specifications was required for production or defense. Only quite recently it became possible to locate two or more individual specifications on correlated subjects bearing the same year of release and publication. This encouraging fact does not necessarily mean that the industry was in a position to produce articles in compliance with those unified requirements of comparable effectivity or was even capable of retooling long established manufacturing lines and processes. In fact, the opposite seems to be the rule rather than exception. Consequently, it is not feasible to furnish an adjusted analysis of the situation. The results of this dissertation depend upon the accuracy, validity, and mutual applicability of the respective source materials and may serve as guides toward further advancement of the standardization of each component part involved. This will provide the background for the co-ordination through specifications that will conform with state of the art and will satisfy the requirements of the users.

The scope of this discussion is limited to current-carrying capacities in varying amounts to 50 amperes of both direct current and alternating current, the latter at frequencies of 60 and 400 cycles per second. The consideration of only copper wires constitutes a further limitation of this paper but the principles and opinions set forth herein lend themselves for interpretation of related conductor and protection characteristics.

Conductors and Conductor Standardization Aspects

The standardization of conductors is basically expressed by recognized wire gauge sizes. In addition thereto, a standardization of permissible conductor loads appears to be desirable. However, the latter depends upon such a plurality of unrelated factors that the preparation of a generally applicable information may be impracticable. Variations in the number of conductors used within a specific run, the material properties of the conductor insulations involved, the load factors of the various conductors, and the type of service such as continuous, intermittent, or short-time duty, add further variables to the elementary considerations. Numerous publications

Paper 58-836, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing May 2, 1958.

R. STEINER is with Radioplane, Division of Northrop Aircraft, Inc., Van Nuys, Calif.

are available on specific conductor arrangements, loadings, and duty cycles to furnish information on specialized applications.[1-4] Those publications represent considerable contributions to the state of the art and extraordinarily skillful applications of basic mathematics to specific, complex electrothermal problems. However, it should be noted that each of these articles was written on the premises of a group of assumptions to simplify the real situation that exists under an admittedly large number of variables. Consequently, it still remains the responsibility of the system designer to appraise each system on its own merits and that usually without the benefit of favorably limiting assumptions.

The leading conductor specifications having official character appear to support this point of view.[5,6] They differentiate, for purposes of wire selection and circuit design, only between a single conductor in free air and cables in conduit and bundles. To concentrate on the elementary portion of the subject matter, only single copper conductors, in free air, manufactured in compliance with the requirements of the applicable specifications, will be considered.[5-7] To clarify the nomenclature, the term "wire" will be used throughout for the description of the respective conductors.

Although the standardized wire gauge system obviates any further discussion, it may be in order to reiterate briefly specific wire properties for future cross-reference. The respective details are enumerated in Appendix I.

The current-carrying capacity of a wire is then primarily a function of the heat transmission or heat balance among the wire and its environment or vice versa. The fundamental equation for the quantity of heat generated by electric current and transmitted by a combination of the heat transferred by conduction, convection, and radiation may be written with practical approximation as follows:

$$0.24I^2Rdt = cWdT + C_{BC}S(T - T_0)dt \quad \text{gram-calories} \quad (1)$$

Expressed in common language, it indicates that the electrical heat input equals the internally stored energy plus the heat loss. The notations to this equation and derivations of subsequent equations are contained in Appendix II.

To focus on the elementary problem, only the case of continuous duty will be considered. It is obvious that the temperature T now stands for the maximum temperature that can be attained as the steady-state heat transmission exists. Consequently, equation 1 converts as follows:

$$0.24I^2R = C_{BC}ST_{max} \quad \text{gram-calories} \quad (2A)$$

Through various substitutions additional equations may be developed from equation 1, such as those for short-time and other noncontinuous duty cycles. However, this would not lie within the contemplated scope.

Of particular interest to the subject discussions is the equation establishing the ratio between the current density of a wire and its diameter, whereby it is necessarily assumed that all wires must not exceed a specified highest temperature. That temperature may vary from one group of considerations to another, provided it remains the same for any one group. Based on the notations given in Fig. 1 and the derivation offered in Appendix II, the afore-mentioned ratio is as follows:

$$\left(\frac{I}{a}\right)^2 = \frac{C_0^2}{d} \quad (5)$$

The result of this calculation indicates that the square of the allowable current density I/a decreases as the wire diameter (d) increases, provided the same ultimate wire temperature is attained in each case.

This phenomenon accounts for the unattractive appearance of the commercial and military cable current ratings which follow only occasionally the conventional increments of the decimal number system.[5-7]

A more significant consequence of the result of equation 5 amounts to the recommendation and requirement, respectively, of both commercial and military wiring specifications that, starting with a wire having a specific, large cross-sectional area multiple conductors shall be used.[6,7] This practice is justified because of the subsequent weight and cost reduction and the increased ease of installation and maintenance. Military specifications set the limit for currents larger than those permissible for copper wire size *AN-2* (Army-Navy). There is obviously no objection to the establishment of a smaller current limit for the requirement of multiconductor applications. Cases are known which have a maximum allowable current-carrying capacity of 10 amp (amperes) for any single wire throughout the air-borne electric system. The advantages attained by compliance with this principle may become more apparent through the presentation of an example: A single copper wire size *AN-2* is capable of carrying current not to exceed 181 amp. The weight of such a wire is 278 pounds per 1,000 feet. If the load of 181 amp is divided between two wires of equal cross-sectional area, wire size *AN-6* needs to be selected, which has a weight of 115 pounds per 1,000 feet per run, or a total material

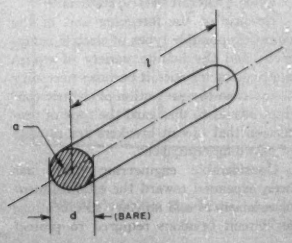


Fig. 1 (left). Wire dimensions schematic

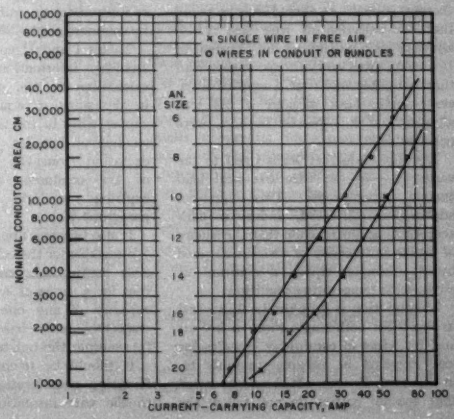


Fig. 2 (right). AN-rated wire current-carrying capacities

Table I. Determination of Conductor Constant for Single Wires in Free Air

Wire Gauge Designation (Fig. 1), AN-Size (1)	Nominal Conductor Area (Fig. 1), CM* (2)	Current-Carrying Capacity (Fig. 2), Single Wire in Free Air, Amps (3)	Current Density (Calculated), Amps/CM (4)	Square of Current Density (Calculated), (Amps/CM)[1] (5)	Diameter of Bare Conductor (Fig. 1), Mils (6)	Conductor Constant (Column 5 × Column 6) (7)
AN-22	755				33	
AN-20	1,200	11	0.00918	0.0000842	41	0.00345
AN-18	1,909	16	0.00840	0.0000706	52	0.00366
AN-16	2,409	22	0.00914	0.0000836	61	0.00510
AN-14	3,830	32	0.00835	0.0000698	76	0.00531
AN-12	6,088	41	0.00675	0.0000456	96	0.00437
AN-10	10,443	55	0.00526	0.0000276	122	0.00337
AN-8	16,864	73	0.00434	0.0000188	167	0.00314
AN-6	26,813	101	0.00376	0.0000141	218	0.00308
AN-4	42,613	135	0.00316	0.0000100	272	0.00272
AN-2	66,832	181	0.00276	0.0000076	345	0.00258
					Average	0.00368

* CM = circular mils.

Table II. Determination of Conductor Constant for Wires in Conduit or Bundles

Wire Gauge Designation (Fig. 1), AN-Size (1)	Nominal Conductor Area (Fig. 1), CM* (2)	Current-Carrying Capacity (Fig. 2), Cables in Conduit or Bundles, Amps (3)	Current Density (Calculated), Amps/CM (4)	Square of Current Density (Calculated), (Amps/CM)[2] (5)	Diameter of Bare Conductor (Fig. 1), Mils (6)	Conductor Constant (Column 5 × Column 6) (7)
AN-22	755	5	0.00652	0.0000425	33	0.00140
AN-20	1,200	7.5	0.00625	0.0000390	41	0.00160
AN-18	1,909	10	0.00524	0.0000274	52	0.00143
AN-16	2,409	13	0.00540	0.0000292	61	0.00178
AN-14	3,830	17	0.00444	0.0000194	76	0.00147
AN-12	6,088	23	0.00378	0.0000143	96	0.00137
AN-10	10,433	33	0.00316	0.0000099	122	0.00126
AN-8	16,864	46	0.00273	0.0000075	167	0.00125
AN-6	26,813	60	0.00223	0.0000050	218	0.00109
AN-4	42,613	80	0.00187	0.0000035	272	0.00095
AN-2	66,832	100	0.00149	0.0000023	345	0.00076
					Average	0.00130

* CM = circular mils.

weight of 230 pounds. The saving in weight, freight, and probably cost is approximately 17% in this case, disregarding the additional payload capacity thereby gained. It should also be noted that the selection of wire size *AN-6*, allowing for a continuous duty current of 101 amp, was very conservative and that a lighter wire with a smaller diameter would have sufficed if its use, allowing for still greater savings, were permissible.[5–7]

Equation 5 further lends itself to the verification of the wire size and load current progressions. This equation, adequately converted, indicates with acceptable approximation that the square of the current density of a wire multiplied by its diameter equals a constant at any specific temperature condition.

This relation is graphically illustrated in Fig. 2 for both single wires in free air and for wires in conduit or bundles for load currents in varying amounts to, in this case, 60 amp. It should be noted that specific entries do not follow the trend of the plot. It appears that this is primarily due to deviations of cross-sectional wire areas accepted by military specifications compared to theoretically more accurate values as adopted by commercial wire tables.[5,8]

Tables I and II contain the data and auxiliary information required for the calculation of the conductor constant in the spirit of equation 5 for single copper wires in free air and for wires in conduit or bundles, respectively. Any deviation from the calculated average of the conductor constant may be explained in light of the preceding paragraph.[5,8]

It also becomes apparent that equation 5 is suitable for the calculation of permissible continuous current-carrying capacities for a series of wire sizes, once the capacity of any one wire size of the respective series has been determined, for example, through test.

It should be re-emphasized that the foregoing considerations do not extend to the wire insulation and its effects. Further, it was not intended to cover short-time ratings of wires, primarily in the absence of the respective military specification requirements, or to discuss the merit of any such stipulations if they were officially provided. It is not obvious as to what advantage would present itself if, for example, it were permitted to subject a specific wire occasionally to 200% of its load for a period of 10 seconds in the course of a flight lasting a period of 10,000 seconds (approximately 3 hours). However, for reasons of cross-reference, mention is made of short-time rating diagrams prepared by nonmilitary sources. For purposes of ready correlation with subsequent considerations, such a diagram was converted to a per-cent load versus time family of curves and is presented in Fig. 3.[7,9]

Conductor Protection

This heading covers a great variety of protection principles, methods, and devices. The treatise awarded these subjects seems to have decreased in inverse proportion to their rising prominence. The operational characteristics of cutouts, fuses, current limiters, relays, automatic switches, and circuit breakers may fall into this category. To keep the following discussion within reasonable limits, it will be concerned only with circuit breakers, particularly with those having an intentional time delay response. The literature available on these topics does not discriminate against or favor any particular means employed to attain time-delay response of circuit breakers. Indications are that thermal circuit breakers allow readily for the calculation of concerted action under specific transient system conditions, whereas empirical methods may be required for the establishment of the respective parameters of magnetic time-delay circuit breakers. It is also noteworthy that the particular literature is concerned, almost exclusively, with circuit breakers used for the protection of wires in aircraft electric systems.[4,6,7,9]

Because of the foregoing and of the countless possible types of electric equipment and the infinite variety of equipment characteristics, it becomes necessary to focus on the protection of electric conductors and to disregard the type of protection that circuit breakers are capable of affording equipment.

Considerable engineering effort has been expended toward the establishment of commercial and military specifications for circuit breakers required to protect

conductors. The salient points of those specifications are the so-called calibration requirements, i.e., the circuit breaker responses from rated current or load through severe short-circuit burdens with respect to the inversely proportional release time periods. The respective military calibration as originally expressed in common language is reproduced partly in Table III and presented graphically in Fig. 4. The latter indicates readily the following:[9,10]

1. The graphical exhibition of requirements, particularly if the current values are expressed in multiples of the rated current, furnishes a perfect impression of the circuit breaker characteristics.[3]

2. The graphically portrayed requirements indicate consistency at current values starting with and larger than 200% of rated load, but an abruptly isolated performance value for the maximum ultimate trip point. This does not permit a true completion of the performance curves.

3. The trend of the consistent portions of the plotted curves compares favorably, however generously, with those of the respective short-time rating diagrams of the wires (see Fig. 3).[7]

4. Through superposition of the diagram shown in Fig. 4 over that presented, in a like scale, in Fig. 3, it becomes evident that the permissible greater tripping time limits of the circuit breakers are considerably smaller than the smallest possible short-time wire-rating time limits. At specific current values a gross differential of 28 seconds exists between permissible wire load and circuit-breaker trip duration. This circumstance invites again questions on the merit of the entire consideration of short-time wire ratings concerning periods of less than, for example, 28 seconds, because it appears that genuine co-ordination among the two elements has been lost.

Table III. Circuit Breaker Detail Calibration Performance Requirements

Test	Circuit Breaker Rating	Per Cent Rated Current	Ambient Temperature C ±2 C	Tripping Time
Minimum ultimate trip	3 -50	115	25	no trip*
Maximum ultimate trip	3 -50	138	25	1 hour, maximum
	3 -50	200	25	13 to 55 seconds
Overload calibration	3 -50	400	25	2 to 7 seconds
	3 -50	600	25	1 to 3.5 seconds
	3 -5	156	−40	no trip*
	3 -5	178	−40	1 hour, maximum
	7.5-50	148	−40	no trip*
	7.5-50	165	−40	1 hour, maximum
Ambient effect on calibration	3 -5	80	+71	no trip*
	3 -5	108	+71	1 hour, maximum
	7.5-50	70	+71	no trip*
	7.5-50	114	+71	1 hour, maximum

* Testing time = 1 hour.

Provided that a reasonable co-ordination between conductors and conductor protection devices can be attained through practical performance specifications and exacting, compatible products furnished by the industry, the establishment of sensible rather than attractive-looking circuit breaker ratings appears to be desirable.

The present relation among permissible current-carrying capacities of wires and circuit breakers is illustrated in Fig. 5. The graphical mode of presentation was chosen to emphasize apparent inequities.[6,10]

The required factual circuit breaker properties are indicated in the bar diagram located in the center of Fig. 5. This diagram furnishes evidence for the merely academic merit of the circuit-breaker rating callouts. For example, the 50-amp circuit breaker does nothing if a current of 50 amp flows through it. It becomes active at a load of 57 amp and must act if subjected to a current of 69 amp, considering various ambient temperature effects. It should be noted that only one wire rating, i.e., 60 amp, is at hand for safe coverage within the 57- to 69-amp operational range of the nominal 50-amp breaker. Comparable situations exist at other circuit-breaker and wire denominations, respectively. Considering the entire, present arrangement within the 0- to 50-amp load range, seven of the 15, or 46.7% of the eligible, specified wire sizes cannot be protected through presently specified circuit breaker ratings. The reason for this startling incompatibility may be the traditional adherence to circuit-breaker rating assignments in the decimal number system or in proportional increments thereof in contrast to the numerically ill-matched, but theoretically correct, progression of wire size designations (see equation 5).[10]

Indications are that in specific areas, particularly outside of the established scope, a somewhat streamlined co-ordination among the respective conductor and its protection device may result. However, it seems to be fair to state that the majority of combinations will present cases of greater imbalance and, consequently, of increased complexity, incompatibility, and lack of standardization between the two factors.

A possible and equitable circuit-breaker rating distribution is offered in Fig. 6 intended to comprise a load current range from 3.45 through 69.0 amp.

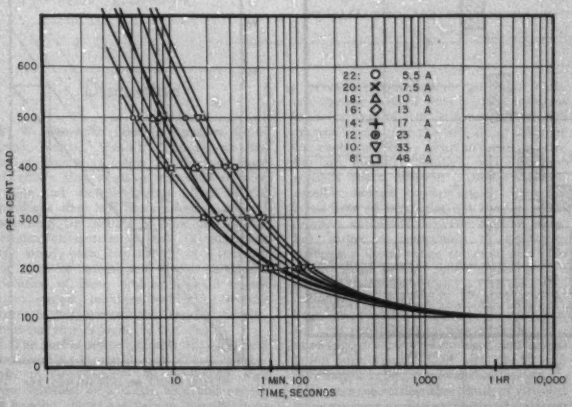


Fig. 3. Aircraft Industries Association short-time ratings, AN-rated copper wires, 3 to 50 amp

Conclusions and Recommendations

The foregoing seems to furnish evidence for the availability of significant results had from substantial pioneering done in the areas of aircraft cable current loads

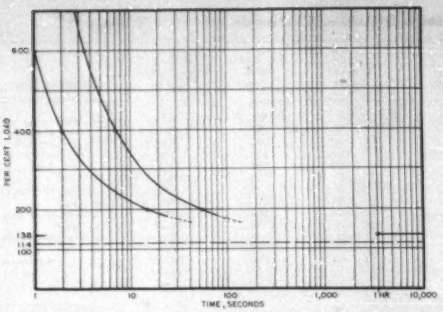


Fig. 4. Performance characteristics of MS25017 (Military Standard) rated circuit breakers

and the correlated temperature variations. Adequate information is on hand on circuit protection devices intended to conform to specific wire characteristics. Both this work and its results have greatly influenced the entire thinking in these two fields. Disregarding the extraordinary and painstaking effort expended and the great variety of cases pursued, the majority of those projects has laboratory status; only a minor portion thereof is immediately applicable to actual design. This is particularly true for circuit protection devices and especially for circuit breakers. It should further be noted that only a small number of those accomplishments has met with official recognition, particularly in military specifications.

It appears that a differentiation is in order between laboratory conclusions and data suitable for field applications. This suggestion refers equally to the practical principles of wire loading, wire size selection, consideration of transients, environmental effects, actually required wire, and circuit protection, and the realistic selection of adequate circuit protection devices. The latter presupposes compatibility among the requirements and properties of the circuit protection devices and those of the respective system wires to satisfy the dependable operation of a completed system design. Such a design will be expected to perform reliably under a combination of most severe electrical and environmental conditions. It is obvious that it will then also operate safely under short-duty and other marginal stipulations.

To alleviate the afore-mentioned situations, it is recommended that the specification requirements for both wires and their protection devices be prepared in a realistic, compatible, and joint fashion.

It is further recommended that new wire protection methods be inaugurated and evaluated, allowing for the equally effective control of hazardous overloads and transient, harmless current stage manifestations without resorting to the indirect time-delay channel and its short-time wire rating implications.

It is also recommended that, in the course of such proposed actions, full consideration should be given to the specific peculiarities of manned aircraft of both the lighter- and heavier-than-air varieties, the pilotless types, and the space vehicles.

Appendix I. Excerpt from the American Wire Gauge (AWG) System

1. The diameter of each (bare) wire size is 112.3% that of the next smaller size.
2. The diameter of every sixth size is approximately doubled compared to the respective smaller size.
3. The cross-sectional area of each size is 126.1% that of the next smaller size.
4. The cross-sectional area of every third

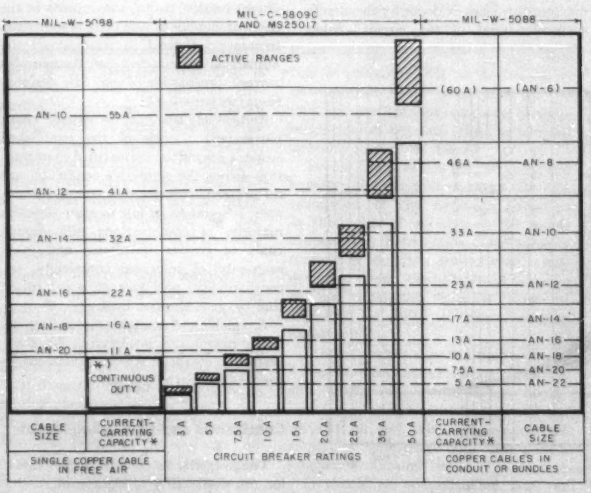


Fig. 5. AN-rated wire and Military Standard rated circuit breaker co-ordination

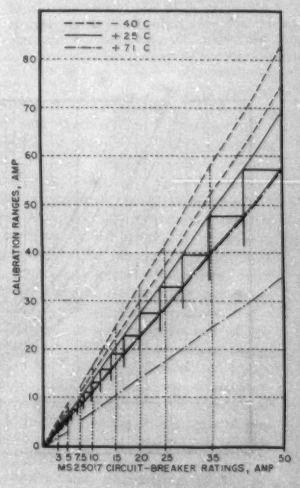


Fig. 6. Proposed equitable circuit-breaker rating distribution, 3 to 69 amp

gauge size is approximately doubled compared to the respective smaller size.

Appendix II. Fundamental Equation for Heat Transmission and Derivation of Specific Equations

The applicable fundamental equation for heat transmission is the following:

$$0.24I^2Rdt = cWdT + C_{RC}S(T-T_0)dt \quad \text{gram-calories} \quad (1)$$

where

I = current, amp
R = wire resistance, ohms
t = time, seconds
c = specific heat, in this case that of copper, gram-calories/degrees centigrade (C), gram
W = weight of the wire portion, grams
T = wire temperature, C (T_0 = the initial or starting temperature)
S = wire surface, square centimeters
C_{RC} = over-all coefficient of heat transfer due to radiation and convection, gram-calories/square centimeters, C, seconds

The element of heat transfer by conduction was considered to be improbable under the respective applications and was disregarded. The temperature rise ($T-T_0$) was entered in the first power because only comparatively moderate temperature changes are involved. For higher temperature levels, the expression for the rise ($T-T_0$) will appear in the 4th power (Stephan-Boltzmann) or in the 5/4 power (Lorenz).

To elaborate on the elementary problem, only the case of continuous duty will be considered. Consequently, equation 1 converts into the following equation indicating a balanced heat transmission state in the absence of stored heat ($cWdt = 0$)

$$0.24I^2R = C_{RC}ST \quad \text{gram-calories} \quad (2)$$

It is obvious that the temperature T is now the maximum temperature that can be attained as the steady-state heat transmission exists. Equation 2 may then be more appropriately expressed as follows:

$$0.24I^2R = C_{RC}ST_{max} \quad \text{gram-calories} \quad (2A)$$

Of particular interest to the subject discussions is the equation establishing the ratio between the current density of a wire and its diameter, which may be derived from equation 2(A), whereby it is assumed that all wires must not exceed a specified highest temperature. This temperature may vary from one group of considerations to another provided it remains the same for any such group.

Through substitution, equation 2(A) may be rewritten as follows:

$$0.24I^2\frac{4l}{K\pi d^2} = C_{RC}\pi dlT_{max} \quad \text{gram-calories} \quad (3)$$

where l = length of the wire portion, centimeters
k = conductivity of the wire, 1/ohms
d = wire diameter, centimeters

Fig. 1 shows the notations used for the foregoing substitution.

Combining the constant values 0.24, 4, π, K, and C_{RC} under a new constant C^3, equation 3 becomes, upon further reduction,

$$I^2 = C^3d^3 \quad (4)$$

By extracting the square root, equation 4 appears in the following useful form:

$$I = C\sqrt{d^3} \quad (4A)$$

For the sake of convenience, equation 4(A) may be rewritten in this fashion:

$$I^2 = C^2\frac{d^4}{d} \quad (4B)$$

With d^4 being a measure for a^2, i.e., the square of the cross-sectional area a, and introducing a new constant C_0, one obtains:

$$I^2 = C_0^2\frac{a^2}{d} \quad (4C)$$

or

$$\frac{I^2}{a^2} = \frac{C_0^2}{d} \quad (4D)$$

Extracting the square root, equation 4(D) appears as follows:

$$\left(\frac{I}{a}\right)^2 = \frac{C_0^2}{d} \quad (5)$$

Equation 5 establishes the ratio between the current density I/a of a bare wire and its diameter (d) for steady-state continuous-duty conditions and a specified maximum temperature (T_{max}).

References

1. Continuous Current and Temperature Rise in Aircraft Cables, M. Schach. *AIEE Transactions*, vol. 71, pt. II, Sept. 1952, pp. 197–203.

2. Continuous Current and Temperature Rise in Bundled Cables for Aircraft, Milton Schach, R. E. Kidwell, Jr. *Ibid.* (Jan. 1953 section), pp. 376–84.

3. D. C. Power Systems for Aircraft, R. H. Kaufmann, H. J. Finison. General Electric Series, John Wiley & Sons, Inc., New York, N. Y., 1952.

4. Aircraft Electrical System Study, H. J. Finison, L. W. Matsch, L. J. Stratten. *WADC Technical Report no. 52-49*, Wright Air Development Center, Wright-Patterson Air Force Base, Dayton, Ohio, Apr. 1952.

5. Aircraft Copper 600-Volt Electrical Wire. *Military Specification MIL-W-5086A*, June 18, 1957.

6. Installation of Aircraft Wiring. *Military Specification MIL-W-5088(ASG)*, June 18, 1956.

7. Design Manual on Aircraft Electrical Installations. Aircraft Industries Association, Washington, D. C., May 1, 1952.

8. Standard Handbook for Electrical Engineers (book), Archer E. Knowlton, editor-in-chief. McGraw-Hill Book Company, Inc., New York, N. Y., eighth edition, 1949.

9. General Specification for Aircraft Trip-Free Circuit Breakers. *Military Specification MIL-C-5809(ASG)*, June 24, 1957.

10. 3 to 50 Ampere Push-Pull Trip-Free Aircraft Circuit Breakers. *Military Standard MS25017*, Jan. 31, 1957.

Discussion

R. W. Stineman (Boeing Airplane Company, Seattle, Wash.): The author has called timely attention to the sad state of affairs that exists in the field of wire rating and wire protection. Two of the problems that face the profession are (1) matching the shapes of wire and circuit breaker characteristics, and (2) matching the steady-state capabilities of a variety of sizes of wires and circuit breakers. Solution of these problems would assure that circuit breakers, while protecting wire, do not unreasonably reduce the current-carrying capacity of the wire.

The author seems to imply that it would be desirable to match each circuit breaker size to two specific wire sizes, one wire size installed as a single wire and one in a bundle. However, various complications not discussed by the author would appear to make the attainment of such an objective a practical impossibility. To begin with, any specific type of wire may be installed in a variety of ambient conditions, the current capacity of the wire being different for each condition. Furthermore, a wire has no unique "bundle" capacity, but actually has a different capacity for each possible bundle configuration. Thus, a given type of wire does not have a set of discrete capacities for the several sizes, but rather a continuum of capacities covering a broad range for each size. No single circuit breaker could match such a range of capacities without severe penalties.

Before attacking wire protection problems, we need to know a lot more about the true capabilities of wire under a wide, unrestricted variety of ambient and installation conditions. The fact that such information is not now available is probably a direct result of the sheer bulk of detailed work necessary using available engineering techniques. However, the recent advent of more sophisticated computing aids may provide the key. I am investigating the possibilities of combining a limited amount of experimental measurements with rigorously executed calculations on a digital computer as a means of obtaining accurate steady-state ratings and time-current characteristics for wire with a minimum of restrictions or assumptions.

With sufficient accurate information on wire characteristics available, it will then be possible to approach the design of more optimum wire protection. It is apparent that there will be a need for much closer control of circuit breaker characteristics, and for a larger number of discrete sizes than is presently available.

R. Steiner: A response to Mr. Stineman's discussion provides the opportunity to acknowledge the concurrence with specific presentations, to interpret particular expressions further, and to complement individual statements of the paper.

Adequate emphasis was placed on the complex union among laboratory and field conditions throughout the paper. It would, indeed, be desirable to attain circuit-breaker characteristics matching with those of wires and the steady-state capabilities of a variety of wire sizes, respectively. It was, however, not intended to intimate that each circuit breaker be capable of monitoring dependably two specific wire sizes. The information in Fig. 5, offered merely to indicate the gaps of protection existing between presently available circuit-breaker ratings and wires both in free air and in bundles, is apparently responsible for this misinterpretation.

Particularly valuable is Mr. Stineman's reference to the effect of ambient conditions on the wire capacities. Although it is not quite apparent as to how any periodically increased wire capacity can be fully utilized by a substantially constant system load, the fact remains that circuit breakers, available now, cannot live up to those broad requirements.

It is encouraging to note that a solution to this intricate problem is sought through advanced computing methods. The answers will undoubtedly present a great challenge to the circuit-breaker industry.

The Limitations of Induction Generators in Constant-Frequency Aircraft Systems

E. ERDELYI
MEMBER AIEE

E. E. KOLATOROWICZ
ASSOCIATE MEMBER AIEE

W. R. MILLER
ASSOCIATE MEMBER AIEE

Synopsis: Starting from basic theory, this paper shows that induction generators are useful only if the speed range is limited. A number of seemingly ingenious schemes are discussed. The conclusion is that there is either no net output or that the efficiencies of the systems are too low for wide speed range operation to find practical application as constant-frequency aircraft electrical supplies.

THE INCREASING demand for electric energy in modern high-performance aircraft has led to the introduction of a-c systems and, in particular, 400-cycle-per-second constant-frequency supplies for the various electronic and accessory power loads. This system is chosen because it can provide large blocks of power with low weight, minimum maintenance, and high reliability. The high efficiency and reliability of the transformers and motors, and the resultant simplification achieved in electronic equipment, make a 3-phase constant-frequency system most desirable.

The mechanical power to drive the alternators can be taken from aircraft gas turbine engines by coupling to an accessory shaft geared to the main shaft or by utilizing bleed air to drive an auxiliary air turbine. Since the main engines are not constant-speed, it is necessary to use some form of infinitely variable transmission if conventional synchronous alternators are to be used. Variable-ratio transmissions are costly, add weight, use valuable space, and require maintenance. Similar observations may be made about accessory air turbine alternator drives. Replacing the conventional alternator and its constant-speed drive with a single electric machine giving a constant-frequency output with a variable-speed input is a worthy goal.

Many system designers have been attracted by the simplicity, low cost, and ruggedness of polyphase induction machines. It is well known that under certain conditions they will operate as generators when driven above synchronous speed. The fact that synchronizing procedure is not necessary when induction generators are paralleled adds even more inducement to designing an induction generator constant-frequency system. This paper shows why even every ingenious induction generator systems fail to provide an adequate constant-frequency power supply for aircraft with engines of wide speed range.

Paper 58-831, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 12, 1958; made available for printing April 10, 1958.

E. ERDELYI, E. E. KOLATOROWICZ, and W. R. MILLER are with the General Electric Company, Erie, Pa.

Analysis of Induction Generator Operation

Physical Description

Fig. 1 is the schematic of an induction motor. The excitation current I_m, drawn from the supply lines, sets up a rotating field with an angular frequency ω_s. The rotor will revolve at a slightly lower speed ω_n. The relative velocity between the field and the stator conductors is ω_s, while the relative velocity between the field and rotor conductors is $\omega_s - \omega_n$. The latter relative speed is generally expressed as $s\omega_s$ where s is the slip defined as

$$s = (\omega_s - \omega_n)/\omega_s \qquad (1)$$

Since ω_s and $s\omega_s$ are in the same direction, the primary and secondary induced voltages will be in phase. For the sake of simplification, the stator and rotor winding configurations are considered identical and iron losses are neglected. The vector diagram in Fig. 2 shows the terminal voltage E, the induced voltage in the primary E_1, the secondary voltage E_2, and the magnetizing current I_m. All the quantities are represented for a single phase. The rotor current lags behind the induced emf (electromagnetic force) E_2 because the rotor circuit is inductive. As in a transformer, the primary winding will draw current from the supply line so as to annul the magnetomotive force of the secondary The sum of the compensating current and the magnetizing current is the primary current. It should be noted that the primary induced emf E_1 and the current I_1 form an obtuse angle θ_1, meaning that the motor is drawing energy from the supply line and transforming it to mechanical energy.

In the case of the induction generator, the angular velocity of the rotor is greater than that of the rotating field. Therefore, s is negative and ω_s and $s\omega_s$ are in opposite directions. The induced emf E_1 and E_2 are in phase opposition as shown in the vector diagram in Fig. 3. The rotor current, however, will still lag behind the emf E_2 and the compen-

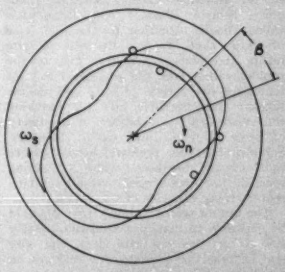


Fig. 1. Schematic diagram of an induction machine

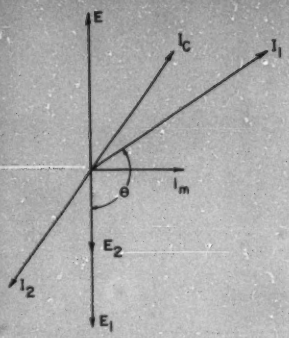


Fig. 2. Vector diagram of an induction motor

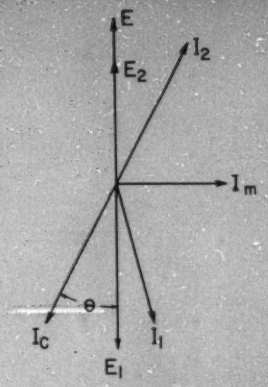


Fig. 3. Vector diagram of an induction generator

sating component of the primary current and the primary induced emf E_1 now form an acute angle. This indicates that energy flow is opposite to that of the motoring mode and the mechanical input to the shaft is converted to electric energy being fed to the supply line.

It is interesting to note that the stator current now leads the primary induced emf. Each value of current is associated with a definite angle of lead. For this reason, the induction generator is unable to provide lagging current as may be demanded by a load. The lagging current required by the load and the magnetizing reactance must be supplied by other sources connected to the line.

Theoretical Analysis

Fig. 1 shows one phase of the stator winding and of the rotor winding. The axes of the two coils are at angle β where

$$\beta=(1-s)\omega_s t \tag{2}$$

The angle β is measured in the direction of the rotating field.

The mutual inductance between a phase of the rotor and stator is X_m when $\beta=0$, i.e., when the two axes coincide. The mutual inductance at time t, assuming $\beta=0$ at $t=0$, is

$$\mathbf{X}_m=X_m e^{j(1-s)\omega t} \tag{3}$$

The voltage equations for the rotor and stator can now be written assuming a short-circuited secondary and using the following notation:

$\mathbf{E}=Ee^{j(\omega t+\theta_1)}$ terminal voltage
$\mathbf{I}_1=I_1e^{j(\omega t+\phi_1)}$ primary current
$\mathbf{I}_2=I_2e^{j(s\omega t+\phi_2)}$ secondary current
X_1 = primary leakage reactance
R_1 = primary resistance
X_2 = secondary leakage reactance at line frequency
R_2 = secondary resistance

$$Ee^{j(\omega t+\phi_1)}=-[R_1+jX_1]I_1e^{j(\omega t+\phi_1)}-jX_m[I_1e^{j(\omega t+\phi_1)}+I_2e^{j(\omega t+\phi_2)}] \tag{4}$$

$$0=-[R_2+jsX_2]I_2e^{j(s\omega t+\phi_2)}-jsX_m[I_1e^{j(s\omega t+\phi_1)}+I_2e^{j(s\omega t+\phi_2)}] \tag{5}$$

Multiplying equation 5 by $e^{j(1-s)\omega t}$ and simplifying equations 4 and 5 gives

$$\mathbf{E}=-(R_1+jX_1)\mathbf{I}_1+\mathbf{E}_1 \tag{6}$$

$$0=-(R_2+jsX_2)\mathbf{I}_2+s\mathbf{E}_1 \tag{7}$$

where

$$E_1=-jX_m[I_1e^{j(\omega t+\phi_1)}+I_2e^{j(\omega t+\phi_2)}] \tag{8}$$

These equations form the basis for the equivalent circuit of an induction machine as shown in Fig. 4.

The quantity $I_2^2R_2$ represents the electric power absorbed in the rotor circuit. The quantity $I_2^2R_2(1-s)/s$ represents the mechanical output. The sum of these two quantities $I_2^2R_2/s$ is the power transferred through the air gap from the stator to the rotor. In an induction motor all of these quantities are positive.

When an induction machine operates as a generator, the quantity s becomes negative. The quantity $I_2^2R_2(1-s)/s$ is now negative too, and the mechanical output

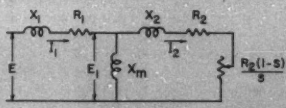


Fig. 4. Equivalent circuit of an induction machine

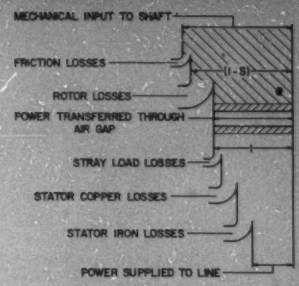


Fig. 5. Power balance of an induction generator

becomes a mechanical input. The rotor loss component remains positive as its expression does not involve the slip. The quantity representing the power through the air gap is negative. The direction of power transfer is now from the rotor to the stator or from the shaft to the rotor and stator electric circuits.

Generator Efficiency

To simplify the discussion, losses in the primary will be neglected; in other words, it will be assumed that all of the power transferred from the rotor to the stator through the air gap goes to the line. The efficiency of this idealized generator is given by

$$\eta=\frac{I_2^2R_2/s}{I_2^2R_2(1-s)/s} \tag{9}$$

or

$$\eta=1/(1-s) \tag{10}$$

Substituting $(\omega_s-\omega_m)/\omega_s$ for s yields

$$\eta=\omega_s/\omega_m \text{ for an ideal machine} \tag{11}$$

For example, the efficiency of an ideal induction generator driven at twice synchronous speed is only 50%. If driven at three times synchronous speed, the efficiency is only 33⅓%. For a nonideal generator, the efficiency figures are reduced as a result of the primary losses. A more detailed power and loss distribution is shown in Fig. 5.

Self-Excitation of Induction Generators

An induction generator may be made self-exciting and will maintain a supply voltage without other sources in parallel if capacitors are used to supply the magnetizing current. The self-excitation is generally initiated by the residual field of the rotor. If C is the capacitance,

the reactance of the capacitor is given by

$$X_c = 1/(2\pi fC) \tag{12}$$

Fig. 6 shows the magnetization curve and schematic connection of an induction machine. The voltage across the capacitor is represented by the line $X_c I_m$. The condition for self-excitation at no load is

$$X_c = X_m + X_1 \tag{13}$$

The reactance $X_m + X_1$ depends on the saturation, and self-excitation can occur only if the capacitance line and magnetization curve intersect. The resultant frequency of operation will be given by

$$f_1 = 1/2\pi\sqrt{C(L_m + L_l)} \tag{14}$$

where L_m and L_l are the primary magnetizing and leakage inductance respectively. This is the resonant frequency of the circuit, consisting of capacitance C and the inductance $(L_m + L_l)$.

If the generator is supplying energy, the output frequency at the stator terminals with a constant speed input will vary as the ratio $1/(1-s)$. The reactance $X_m + X_1$ decreases in the same fashion and the reactance of the capacitor increases by $(1-s)$.

To satisfy the self-excitation condition, the reactance $X_m + X_1$ must increase as X_c increases. Self-excited operation becomes possible only at a less saturated level. The terminal voltage of the induction generator will therefore be decreased.

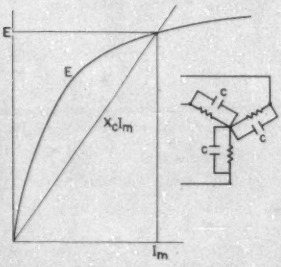


Fig. 6. Self-excitation of an induction generator

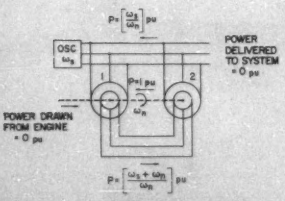


Fig. 7. Two-machine scheme with rotor frequency sum of stator and shaft frequencies

The capacitor must supply the reactive power for the magnetizing reactance, the leakage reactance, and the reactive component of the load. The capacitor, therefore, must be increased with increased loads or with more inductive loads, to keep the voltage constant. Schemes using capacitor switching and saturable reactors exist for automatically keeping the voltage and frequency of induction generators constant.

Discussion of Some Previously Proposed Systems

A number of schemes have been proposed involving two induction generators on a common shaft with interconnected windings to overcome the limitations of single unit systems. These schemes fall into two categories. Either power circulates between both machines and no net power is supplied to the line, or power proportional to slip must be dissipated in the rotor circuitry at severe efficiency penalties.

Systems With No Net Output

The basic system proposed most frequently falls into the first category and is even patented in a Swedish patent[1] and a British patent.[2] It consists of two wound rotor induction machines mounted on the same shaft. A "small" oscillator is proposed to supply the stator of the first machine with a rotating field of such direction that the rotor frequency is the sum or the difference of the stator (ω_s) and rotor (ω_n) frequencies. This "exciter rotor power is applied to the "main" machine in the opposite phase sequence so that the shaft speed frequency component cancels out in the rotating field and hence in the induced stator voltage of the main machine which supplies constant frequency power to the line.

The missing link in this beautiful chain of operations is, of course, the power balance; as demonstrated in the foregoing, the powers divide according to the frequencies. The shaft frequency is ω_n, the line or stator frequency is ω_s and the frequency in the rotor circuit is $(\omega_s \pm \omega_n)$. Assuming the power delivered to the exciter by the shaft to be 1.0 per unit and that the machines are ideal, leads to the following analysis:

Fig. 7 shows the power flow P, when the rotor frequency is the sum of shaft and stator frequencies. The shaft supplies 1.0 per unit mechanical power; ω_s/ω_n per unit power is drawn from the

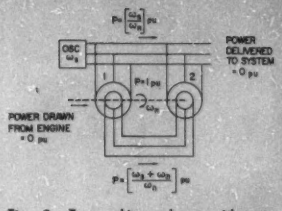


Fig. 8. Two-machine scheme with rotor frequency difference between stator and shaft frequencies

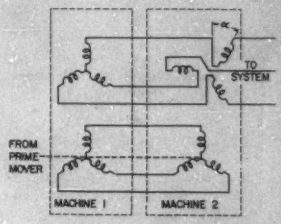


Fig. 9. Two-machine series scheme

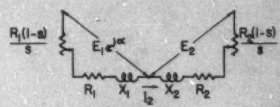


Fig. 10. Equivalent circuit of scheme of Fig. 9

stator, and therefore $[1+(\omega_s/\omega_n)]$ per unit is supplied electrically at the rotor terminals. This total power flows into the rotor of the second machine. Of this, 1.0 per unit must be returned to the shafts by this machine as mechanical power and ω_s/ω_n per unit then flows through the air gap and is available as stator electric power. However, this is exactly the power required to excite the stator of the exciter in the first phase, and no net power is available from the system. The first machine in effect operates as combination transformer and induction generator and the second as a combination transformer and induction motor.

When the rotor frequency is the difference between shaft and line frequencies as in Fig. 8, the exciter acts as a generator delivering ω_s/ω_n per unit to the line and $(\omega_s - \omega_n)/\omega_n$ per unit power to the second machine rotor. To satisfy the balance, this second machine must deliver 1.0 per unit to the shaft and therefore must draw ω_s/ω_n per unit power from the line, the

exact amount the exciter had supplied. The first machine, or exciter, acts as a generator in both rotor and stator circuits and the second machine acts as a motor, running supersynchronously with power supplied to both rotor and stator. Again, no net power is available from the system.

Systems Dissipating Rotor Energy

The other major classification of constant-frequency systems using induction machines most often suggested reduces to two or more machines driven supersynchronously and acting as generators with the rotor power component dissipated as loss.

A system in this classification, described in some length in a British patent,[5] received considerable attention when literature and a wooden model were displayed at the 1957 Paris Air Show. In this system the rotor windings of the two machines are connected together as shown in Fig. 9. The stator windings are connected in series but with winding axes shifted by a fixed angle. By controlling the saturation of one of the stators, the output voltage can then be continuously adjusted. Frequency is held constant by capacitance coupled through an adjustable coupling transformer.

From a power point of view, however, this system is no different from two induction generators connected in series. It can therefore be reduced to one equivalent circuit as shown in Fig. 10. If 1.0 per unit power is drawn from the shaft, ω_i/ω_n per unit $= 1/(1-s)$ per unit is delivered to the stators as electrical output and the remainder $(\omega_n-\omega_r)/\omega_n = -s$ per unit must be dissipated as losses in the rotor circuit.

Conclusions

An induction generator is not suitable as a constant-frequency power supply if it must operate over a wide speed range, because of the large amount of power that must be dissipated at large slips encountered at cruising speed. The necessity for providing a separate source of reactive power is a further complication. Arrangements using additional induction machines fail to overcome the shortcomings.

References

1. Växel Strömsgenerator för Konstant Frekvens och Spänning Oberoende av dess Varvtal, **K. Polasek, K. Jonsson.** *Swedish Patent no. 149,392*, Jan. 20, 1955.

2. Improvements in or Relating to Electrical Generators, **K. Polasek, K. Jonsson.** *British Patent no. 700,036*, Nov. 25, 1953.

3. Improvements in or Relating to Electrical Generators, Asea Material A B. *British Patent no. 773,867*, May 1, 1957.

4. The Nature of Polyphase Induction Machines (book), **P. L. Alger.** John Wiley & Sons, Inc., New York, N. Y., 1951.

5. Theory of Alternating-Current Machinery (book), **A. S. Langsdorf.** McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1955.

6. Capacitive Excitation for Induction Generators, **E. D. Bassett, F. M. Potter.** *AIEE Transactions (Electrical Engineering)*, vol. 54, May 1935, pp. 540–45.

Discussion

L. J. Stratton and K. M. Chirgwin (Jack & Heintz, Inc., Cleveland, Ohio): The authors describe an arrangement employing two induction generators on a common shaft and show that this system has no net output. It is possible to arrive at this same conclusion by considering these two machines as a pair of induction machines connected in cascade.

Fig. 11 shows two cascade-connected machines on the same shaft using the symbolism of Figs. 7 and 8. The basic equation relating the frequency of the output voltage to that of the input voltage and to the speed of rotation of the shaft for any cascade machine pair is

$$\omega_0 = \omega_i + 2\pi n(p_1 \pm p_2) \tag{15}$$

where

ω_0 = frequency of the output, radians per second

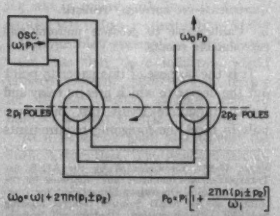


Fig. 11. Two cascade-connected machines on same shaft, with use of symbolism of Figs. 7 and 8

ω_i = frequency of the input, radians per second

n = speed of rotation of the shaft, revolutions per second

p_1 = number of pairs of poles of first machine

p_2 = number of pairs of poles of second machine

The + sign applies to cumulative cascade connection

The − sign applies to differential cascade connection

In the system considered by the authors, the input and output of the cascade machine pair are to be connected together. To do this the input and output frequencies must be equal. Equation 15 shows that these frequencies are equal only when the speed of rotation of the shaft is zero (in which case the apparatus performs as a transformer), or when the quantity $(p_1 \pm p_2)$ is equal to zero. The latter is a case that the authors consider, and it occurs only when a differential cascade connection is used and when the number of pole pairs on the two machines are equal. The fact that this is now a cascade machine pair having an effective number of poles equal to zero should not pass unnoticed.

The authors refer to a fundamental relationship which applies to a single induction machine or to any number of cascaded induction machines, and which states that, if all losses are neglected, then the power in any given part of the system is proportional to the frequency in that part. Applying this relation we see that, by multiplying each term of equation 15 by a constant, we obtain equation 16, which shows the relationships between the powers in the various parts of the system:

$$P_0 = P_i\left[1 + \frac{2\pi n}{\omega_i}(p_1 \pm p_2)\right] \tag{16}$$

We have already seen that, under the conditions assumed, the input and output frequencies must be the same and that $(p_1 \pm p_2)$ is equal to zero; it therefore follows that the input and output powers must also be equal, or the output from the system is equal to the input to the system from the oscillator, and the power transferred from the shaft to the system is zero. This confirms the authors' findings.

The authors conclude that an induction generator is not suitable as a constant frequency power supply if it must operate over a wide speed range. We believe this statement is intended to apply either to a single squirrel-cage induction generator or to a cascade combination of two induction generators, in which case we entirely agree with the authors' findings. However, to avoid any misunderstanding, it should be pointed out that a wound rotor induction generator used in combination with other auxiliary (commutator type) machines enables one to build a constant-frequency variable-speed system which does not require dissipation of large amounts of slip power. These schemes are already well known as constant-frequency variable-speed motors by such names as Scherbius and Krämer. Any of these schemes can be modified to operate as generating systems. The reason they have not been adopted on airplanes is not because of the induction generator itself, but because of the auxiliary machines which must be used in conjunction with it.

In fact, it is quite probable that an all-electric constant-frequency variable-speed generating system for aircraft will include an induction generator.

Morris Flugstad (Boeing Airplane Company, Seattle, Wash.): This timely paper should serve to minimize some of the differences of opinion on the relative

merits of various variable-speed constant-frequency a-c systems employing induction machines as generators.

Engineers in the adjustable-speed induction motor field have been faced for many years with the analogous problem of separating the bad from the good. For example, some time ago, the writer had occasion to evaluate various adjustable speed induction motor schemes. One of the schemes included a tandem machine similar to one described in this paper. With equivalent circuits and circle diagrams it was shown that the system was equivalent to a single induction motor operating on variable stator voltage for speed adjustment. Such a system obviously dissipates power in the rotor circuits proportionally to slip speed and shaft torque and as such is very inefficient when employed in wide-speed-range application, as is indicated in the authors' analysis.

However, many induction motor schemes employ efficient schemes for transmission of slip power and as such are efficient when used over a wide-speed range.

The authors make a serious omission in neglecting inclusion of the generator counterpart to the classification of adjustable speed induction motors utilizing arrangements whereby slip power is efficiently transmitted. The mode of operation of an induction machine employed as a generator, where slip power is salvaged, falls strictly into the induction generator category. The authors' conclusion that induction generators are useful only if the speed range is limited, is correct only for the special class of machines discussed in this paper. Induction generators employing means of slip power transmission generally utilize a wound rotor induction generator as the main unit and various other rotating and/or static electric components to effect efficient handling of slip power. Some arrangements falling in this classification now show great promise as efficient variable-speed constant-frequency generators.

E. Erdelyi, E. E. Kolatorowicz, and **W. R. Miller:** The analysis given in the Stratton-Chirgwin discussion confirms the obvious conclusions reached in the paper. Mr. Flugstad points out very correctly the timeliness of the paper and the authors are gratified by the general agreement about the limitations of induction generators.

All three discussers deplore the authors' omission of the generator counterparts of adjustable-speed induction motor drives using auxiliary machines such as the Krämer and Scherbius systems. These systems, of course, can salvage the slip power. It was the authors' intent to point out the efficiency limitations of cascade systems which have been repeatedly offered as solutions to the variable-speed constant-frequency problem.

As pointed out by Stratton and Chirgwin, induction generator systems using auxiliary machines are not attractive for aircraft systems. If conventional rotating auxiliary machines are used, they require commutators and slip rings and would be quite unsatisfactory at the high slip frequencies and powers involved. Even assuming that power would flow in both directions in the auxiliary machines, they would be quite large for a wide-range system. In addition, the wound rotor induction generator tends to be substantially heavier than a salient pole synchronous machine of the same rating.

Static components could be used in lieu of rotating machines. To minimize the maximum power flow, the conversion device should be capable of transmitting power in both directions. This bidirectional, variable-frequency device would be quite complex. By transmitting slip power in one direction only, simple circuitry can be used, but the power handled would be increased considerably. It becomes questionable whether it would not be better to convert the output of a variable-frequency generator directly. At present, suitably rated static components are not available but this situation may change in the near future.

It is hoped that continued effort will lead to the design of a reliable, stable, constant-frequency aircraft power supply of reasonable weight.

Aircraft Switch and Relay Inductive-Circuit Test Loads

J. P. DALLAS
MEMBER AIEE

Synopsis: Present contact-system inductive-circuit test procedure and standard inductors are shown to be inadequate for evaluating the performance of aircraft and missile relays and switches. The need for inductive-circuit testing of contact systems is shown to be critical since contact deterioration is proportional to the energy dissipated in the contact gap which, for inductive circuits, is likely to be many times the energy dissipated in interrupting resistive loads. An inductive test procedure for contact systems and a standard inductor design are proposed which can closely simulate aircraft and missile inductive loads and are calibratable to a practical degree.

OVER the years some of the fundamentals, affecting the life of relay and switch contact systems, have been obscured—buried under avalanches of specification compromise. Basic principles have been neglected and compromise has been applied to compromise through each generation of specifications. From the 1940 U. S. Air Force and U. S. Navy specifications down to present military specifications, contact-system tests have been insufficiently related to the basic physics of electric circuit operation.

When the lack of correlation between switch and relay inductive-circuit test procedure and electric circuit operation is mentioned, present practice is usually justified on the basis that, "We're getting by all right and we don't have significant inductive-circuit interruption failure." It is true beyond question that if the industry is satisfied with present levels of reliability and cost there is no reason to change. However, if substantially higher levels of contact-system reliability are desired then more realistic contact-system test procedure is likely to be worth the effort. Present unrealistic test procedure leads to unnecessary expense in design and production, but this cost is difficult to evaluate. Were such an evaluation made, a startling increase in interest in this subject could be predicted.

To be satisfactory, a contact-system inductive-load test procedure must be:

1. The simplest method which will surely evaluate performance and reliability.
2. Realistic; it should closely simulate actual usage conditions. Aircraft switches and relays are commonly designed to meet military specification test requirements rather than specific air vehicle usage requirements. The whole industry design and production can, therefore, expensively drift away from real needs unless test requirements are currently realistic.
3. Calibratable to produce uniform and reproducible results.

It is the purpose of this paper to point out the degree to which present relay and switch inductive-circuit test procedure fails to meet the foregoing requirements

Paper 58-841, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 18, 1958; made available for printing April 16, 1958.

J. P. DALLAS is with the Hughes Tool Company, Culver City, Calif.

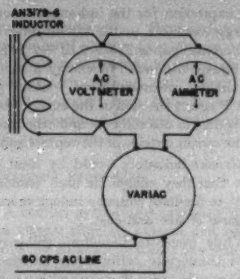


Fig. 1. Inductance computed from volt-ampere measurement at 60 cycles per second with rated *rms* current

Fig. 2. Inductance measured with standard inductance bridge

and to suggest a more realistic and calibratable approach to this problem.

History

In 1941 it became apparent that some method of simulating aircraft inductive-circuit switch loads was necessary for proper testing of relay and switch contact systems. In 1942 the Bureau of Aeronautics, Washington, D. C., initiated a study of inductive-circuit switching loads common in aircraft electric systems. Gun-firing-solenoid circuits were found to be the most critically inductive loads. It was concluded that circuit inductivity could be described by the time constant of the circuit expressed as the ratio of the inductance L divided by the resistance R. The L/R value of the particular gun solenoids in use was stated to be 0.026 second. It was thought that if contact systems were tested for an inductive-circuit load equivalent to these gun-firing solenoids they would be adequate for all other applications. A series of five standard reactors for loads ranging from 5 to 100 amperes were designed.

In 1943 the *AN3179*[1] standard test inductor drawing was issued. This drawing was in accord with Bureau of Aeronautics test inductor specification *E180*.[2] The unjustified assumption, as is shown later, was that these inductors would provide uniform inductive-load testing for all d-c contact systems used in military aircraft electric equipment.

In a paper by Quill and Rader[3] the authors state that "The standard test inductors specified by the Bureau of Aeronautics, Navy Department, *E180* [*AN-3179*] saturate at about 20 per cent of their rated current. This reactor will then have an appreciable inductance up to 20 per cent of its rating and negligible inductance thereafter." The inductors described in Bureau of Aeronautics drawing *E180* and in the *AN3179* Standard specify the core air-gap dimension. It has been stated by the manufacturer that changing this air-gap dimension from the figure specified will provide an adjustment which will allow unsaturated operation of the cores. To do this, of course, will reduce the inductance to a fraction of the 0.026 L/R ratio specified in *MIL-R-6106C* (Aircraft Standards Group) aircraft relay general specification.[4]

From the data presented in a paper by Rader and Litscher,[5] also in 1944, it is evident that L, the inductance, of an iron-cored coil cannot be exactly defined or measured to values close enough for standard and/or test use. Values for the same inductance are shown to vary as much as 300% depending on which standard method of measurement is used and/or the magnetic history of the inductance. It is suggested that anyone considering this problem would do well to start by a careful study of this excellent discussion.[5]

Some of the conclusions of Rader and Litscher may be very easily experimentally verified in the laboratory. If the inductance of an *AN3179-6* 6-ampere test inductor standard is measured by the four standard methods shown in Figs. 1 through 4, five very different values ranging from approximately 0.13 to 0.009 henry may be obtained. The word approximately is used advisedly because any two of these inductors are likely to show as much as 100% difference in the measured values of L for the same test method, such as shown in Figs. 1, 3, and 4, where the core will be saturated as it will be in test usage.

The current military specifications for relays and switches, such as *MIL-R-6742*,[6] *MIL-S-3950*,[7] and *MIL-R-6106C*, call for inductive-circuit contact-system tests using the previously described *AN3179* standard inductance and, in the case of specification *MIL-R-6106C*, general specification for aircraft relays, require that the time constant of the test circuit using these inductors shall be 0.026 second; an obvious impossibility as pointed out in reference 3.

Need for Contact-System Inductive-Circuit Test

The major factor determining the life of current-carrying contact systems is the contact metal oxidation, erosion, and transfer resulting from electric discharge in the contact gap. In general, contact-system deterioration is proportional to the energy dissipated in the contact gap. In a typical inductive circuit controlling a relay or solenoid coil, the energy discharged in the contact gap on circuit interruption is likely to be many times the energy discharged from interrupting a

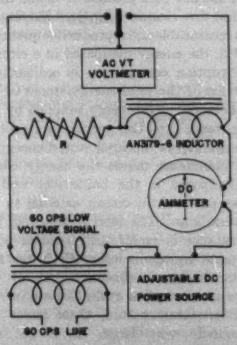


Fig. 3. Inductance computed from 60-cycle-per-second impedance while loaded with rated direct current

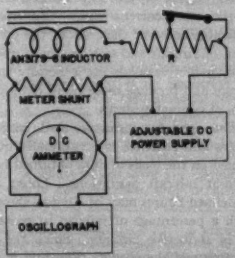


Fig. 4. Inductance computed from oscillograph recording of time constant for incremental decreases of current

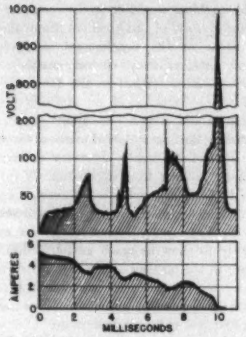


Fig. 5. Typical 5-ampere inductive-circuit-interruption contact-gap energy

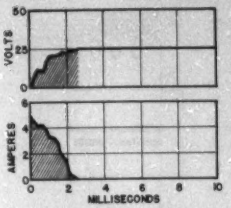


Fig. 6. Typical 5-ampere resistive-circuit-interruption contact-gap energy

normal resistive circuit of the same current. Fig. 5 is drawn from typical aircraft fire-extinguisher solenoid-circuit-interruption oscillograms. Integrating the values of current and voltage shown in the figure for the period of circuit-interruption arc time shown, 2.69 joules of energy dissipated in the contact gap are obtained. By referring to Fig. 6, however, it is seen that an equivalent 5-ampere resistive-circuit interruption is accompanied by a dissipation of less than 0.2 joule. Note that the energy dissipated in the inductive circuit is of the order of ten times greater than that for a resistive circuit handling the same current. It is not intended to imply that inductive-circuit interruption is likely to produce ten times the contact wear that an equivalent resistive circuit would. Considerable contact wear in resistive circuits occurs during the closing cycle of operation. With inductive circuits the rate of current rise determined by the time constant is such that the contacts are normally completely closed several milliseconds ahead of the rated current value. Experience indicates however that, under the conditions of aircraft and missile circuit operation, inductive loads are generally more severe and produce more rapid deterioration of contact systems.

The need for inductive-load tests of relays and switches is occasionally questioned on the basis of data obtained from current aircraft operations. Loose and ill-defined figures may be cited to show as high a percentage of failures in resistive as in inductive circuits. Since the inductivity of the circuits discussed is not defined and account is not taken of cycles of operation and many other critical factors, such data have little significance. One might draw the conclusion with some logic that present military specification contact-system test requirements are so far from reality that the results obtained from such tests and the performance of equipment designed to meet such test requirements may be irrelevant.

Fundamental Considerations

An inductive-circuit test procedure should relate the energy from the test circuit dissipated in the contact gap to the maximum contact-gap energy found in anticipated applications of the switch or relay. The first step then in creating a logical inductive-circuit test procedure is to determine the joules of circuit-interruption contact-gap energy for typical aircraft and missile circuits judged to be critical.

The energy stored in an inductance is usually defined as joules $= 1/2\ LI^2$, where L equals the inductance of the coil and I equals the current in amperes. This expression is not directly useful in computing the energy dissipated in a circuit-interruption contact gap for two reasons: 1. it has not been possible to define the L of iron-core electric equipment inductance as a measurable and reproducible quantity, and 2. the energy dissipated in a circuit-interruption contact gap is ordinarily a fraction of the total energy stored in the inductance. The energy supplied by the inductance to the circuit-interruption contact gap is the total stored energy of the inductance minus the energy which is dissipated in the inductance and in other parts of the circuit external to the contact gap. The stored energy of the inductance is partly dissipated in eddy currents, hysteresis losses, I^2R losses, and dielectric losses. Because part of the circuit-interruption cycle and arc is of a high-frequency nature, these losses are commonly very large. It is not uncommon for a solid-core solenoid to dissipate 80% of its theoretical stored energy in such losses. One could, then, rewrite the expression for the inductance stored energy available to the circuit-interruption contact gap as $J_c = 1/2\ LI^2 - P$, where J_c equals the joules of energy available to the circuit-interruption contact gap and P equals the sum of the energy losses dissipated in the coil core and other parts of the circuit outside of the contact gap.

Air-core inductors would be ideal except that they cannot be used economically to simulate common inductive loads because of size and cost.

While iron-core inductors will never yield completely calibratable standards, steps may be taken to ensure within useful tolerances the uniformity of inductance values. To do this, the following limitations must be observed: 1. The inductor must be operated substantially below saturation of the iron core. Iron-core inductors operated near saturation do not have stable values of inductance. 2. Economic design will result in large air gaps. 3. The test inductor must be operated at flux densities above some minimum value which must be determined for each core assembly. 4. Adjustment of inductance values by air gap or magnetic shunts is not satisfactory. Use a tapped coil or multiple coils. 5. A contact-system test inductor should be calibrated by plotting differential inductance per turn against ampere-turns. This may be accomplished from an oscillographic recording of small changes in the time constant. The ampere-turns must always be varied from maximum value to smaller values in order to avoid major hysteresis-loop inaccuracies. The opening of switches shunting sections of a series resistor can

Fig. 7. Test inductor

be used conveniently. From such a curve the range of ampere-turn change which will yield a reasonably constant inductance can be selected. A typical test-inductor design may have a usefully stable inductance between 1,000 and 10,000 ampere-turns.

Proposed Procedure

A contact-system test procedure which can closely simulate aircraft and missile inductive loads calibratable to a practical degree is proposed as follows:

1. Joules of circuit-interruption contact-gap energy for typical aircraft and missile inductive circuits judged to be critical are computed from oscillograms of circuit-interruption contact-gap currents and voltage as illustrated in Fig. 5. Integration of the voltage and current curves illustrated in Fig. 5 will give joules of contact-gap dissipation energy which has previously been represented as J_r to distinguish it from the total stored energy J of the inductance.

2. Use an inductor constructed with sufficient air gaps and operated at flux densities that have been determined to maintain a nearly constant incremental inductance throughout its range of adjustment. The inductor should be tapped to provide close adjustment. It is useful to determine the core constant, in microhenrys, for a turn. Since inductance varies as the square of the turns, the required number of turns to be used for a given inductance is $N=\sqrt{L/K}$ where N is the number of required turns, L is the required inductance in microhenrys, and K is the core constant in microhenrys.

3. In a test circuit the inductor should be adjusted to a value which will produce in the contact gap of the test sample the joules of circuit-interruption energy specified. The specified value of J_c is to be the result of the determinations discussed in item 1. The adjustment of the inductor may be checked with a dual-beam oscilloscope equipped with a camera.

4. For an approximate adjustment of the test inductor, determine the equivalent effective inductance L_K for producing the specified energy in the contact gap by substituting the appropriate values of J_c and I in the expression $L_K=2J_c/I^2$ where L_K is henrys of inductance, J_c is the specified contact-gap circuit-interruption energy in joules, and I is the current in amperes.

Fig. 7 shows an inductor as described in the foregoing which has been in use for inductive-circuit contact-system testing for the last 8 years. A 3- by 4-inch-section laminated-transformer-grade silicon-steel core is used with a 1/2-inch core gap, with coils for 0.006 henry to 11 henrys and currents from 4 amperes to 400 amperes. The coils are heavily insulated to take a 5,000-volt high-potential test and are provided with protective safety gaps on the front of the connection panel. A core constant of 1.77 microhenrys can be maintained for a range of 2,000 to 12,000 ampere-turns with a 1/2-inch gap in the core.

References

1. Standard for Test Inductor. *Army-Navy Standard AN 3179*, 1943.

2. Specification for Test Inductors. *U. S. Navy Specification E180*, Bureau of Aeronautics, Navy Department, Washington, D. C., Feb. 1943.

3. D-C Arc Interruption for Aircraft, J. S. Quill, L. T. Rader. *AIEE Transactions*, vol. 63, Dec. 1944, pp. 883–88; *discussion*, pp. 1472–73.

4. Aircraft Relays. *Military Specification MIL-R-6106C (ASG)*, Aircraft Standards Group, U. S. Air Force, May 15, 1957.

5. Some Aspects of Inductance When Iron Is Present, L. T. Rader, E. C. Litscher. *AIEE Transactions*, vol. 63, Mar. 1944, pp. 133–39.

6. General Specification for 115/200 Volts AC or 120 Volts DC Aircraft Relays. *Military Specification MIL-R-6742*, U. S. Air Force, June 14, 1950, May 17, 1955.

7. Toggle Switches. *Military Specification MIL-S-3950*, May 17, 1955.

8. Comments on Aircraft Switch Testing, T. R. Stuntpnagel, J. P. Dallas. *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 301–05.

A Discussion of D-C High-Potential Test Voltage for Aircraft Electrical Insulation

L. B. KILMAN
MEMBER AIEE

J. P. DALLAS
MEMBER AIEE

Synopsis: D-c high-potential test voltage that will detect aircraft equipment insulation flaws in a manner equivalent to the 60-cycle rms high-potential test voltage now specified is investigated. The purpose of equipment high-potential tests is stated to be the detection of insulation faults which are shown to be of two categories: 1. holes, cracks, discontinuities, and all faults which result in small-hole air gaps in insulation; and 2. marginally thin insulation, and insulation where the effective thickness has been reduced by the inclusion of foreign matter. A d-c/a-c rms test-voltage ratio of 1.41 is established for the detection of small-hole defects in insulation. Because insulation materials of the type used in aircraft equipment are not available in sufficiently thin thicknesses for breakdown tests at the currently used 1,500-volt 60-cycle rms high-potential test voltage, conclusive data could not be obtained regarding the d-c/a-c high-potential test-voltage ratio necessary to detect marginally thin insulation. A d-c/a-c high-potential test-voltage ratio of 1.6 is tentatively suggested for detecting marginally thin insulation pending further investigation.

THE QUESTION of what d-c high-potential test voltage is equivalent to a standard 60-cycle a-c high-potential test voltage has been under consideration for more than 30 years and much vigorous discussion has failed to generate any general agreement on this question. D-c high-potential voltage from 1.2 to 2.4 times the standard 1-minute 60-cycle rms test voltage commonly specified have been advocated.[1–4] The reason this subject has arisen so often is that d-c high-potential testing has many advantages as well as serious disadvantages. The advantages stated for high-potential d-c testing are as follows:

1. In testing large electric machinery a d-c testing potential avoids quadrature currents from internal capacitance. Substantial alternating currents not concerned with insulation breakdown make it difficult to detect actual insulation faults.

2. D-c high-potential test equipment for large electric machinery may be safer to use due to the fact that the current output of the test unit can be limited to a relatively small value compared with a-c test equipment.

3. It is asserted that an approach close to the breakdown point of the insulation can be predicted in d-c high-potential testing from changes in the leakage current. It may be noted that there is some doubt that this will apply to equipment insulated with the better aircraft insulations now in use, some of which do not show any significant increase in leakage current prior to rupture.

Paper 58-945, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, June 22–27, 1958. Manuscript submitted March 18, 1958; made available for printing April 16, 1958.

L. B. Kilman and J. P. Dallas are with the Hughes Tool Company, Culver City, Calif.

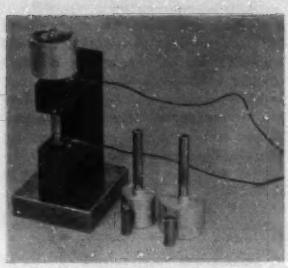


Fig. 1. Test jig for testing breakdown voltage

4. Dielectric stress with direct current is more evenly distributed.

The disadvantages of d-c high-potential testing are:

1. Primarily, no agreement has been reached on the voltage to be used which will be equivalent to standard 60-cycle a-c high-potential tests.

2. D-c high-potential testing may not simulate as closely, as the a-c tests do, critical operating stresses such as circuit-interruption transients. Under high-potential a-c test, the insulation is subject to vibrating mechanical forces resulting from the action of changing magnetic and capacitance fields.

3. In large electric machinery the time required for the d-c test voltage to stabilize due to dielectric absorbtion currents when the test potential is first applied may require an increase in the time for the test.

4. In general, d-c test equipment will have a cost and size disadvantage for use in testing aircraft electrical items of less than 30 kw.

5. With d-c high-potential testing a dangerous charge may remain on the equipment and provisions for its discharge are necessary in some cases.

Part of the controversy over d-c high-potential test voltages may exist because there is a tendency in such discussions not to completely define the objectives under consideration. This discussion is concerned exclusively with high-potential tests of electric equipment insulation for the purpose of detecting insulation flaws, holes, cracks, discontinuities, and inclusions. Dielectric-strength tests as applied to testing insulation material are not considered. The entire purpose of equipment dielectric tests should be to discover flaws in the insulation or marginally thin insulation which must be detected because inductive-circuit-interruption transient voltages in aircraft circuits may reach values of more than 1,000 volts in aircraft equipment.[6–10] If these discontinuities and/or insulation faults are not detected by high-potential tests, then in operation, inductive-circuit-interruption transients may be discharged through these flaws or may puncture marginally thin insulation and cause insulation failure.

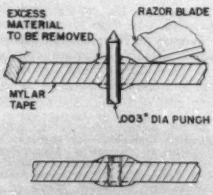


Fig. 2. Typical punched or drilled hole and method of removing burr or excess material

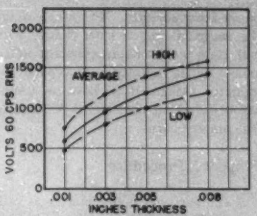


Fig. 3. A-c rms high-potential breakdown voltage versus insulation thickness through 0.003-inch-diameter holes in Mylar. Ten tests made at each indicated thickness

—— Average breakdown voltage
- - - Maximum and minimum breakdown voltage

Method of Determining Equivalent D-C High-Potential Test Voltages

In determining a d-c high-potential voltage test suitably equivalent to present 60-cycle rms test-voltage requirements, it is convenient to divide equipment insulation flaws into two categories as described in the synopsis. The reason for making this division of the problem is that the breakdown of air gaps and the breakdown of solid insulation may not occur at the same ratios of direct to alternating voltage.

The major part of this discussion is concerned with the small-hole air-gap type of insulation faults rather than with marginally thin insulation faults because there has already been considerable investigation of solid insulation.[1–5]

For purposes of this investigation a test jig shown in Fig. 1 was constructed in accordance with ASTM *D149-55T* specification for the test of insulating tapes.[11] Cylindrical tungsten 1/4-inch-diameter electrodes with edges rounded to a radius of 1/32 inch and with the upper electrode weighted to rest on the insulation with a force of 1/10 pound were used. The electrode surfaces were ground square and lapped to a 10-microinch finish. They were repolished after each discharge.

Holes 0.003 inch in diameter were punched in Mylar tape insulation of 0.001-, 0.003-, 0.005-, and 0.008-inch thickness. The technique of punching uniform-diameter holes without thickening the insulation around the hole in materials which tend to cold flow, such as Mylar, is a real problem. Various drilling methods and standard punch-and-die techniques were tried without success. Fig. 2 shows a microscopic section of a typical piece of Mylar with a typical hole. When pressure is applied to such a hole, either in measuring the insulation with a micrometer or in the test jig, the raised portion of material around the hole flows back into the hole, tending to close it up, of course, such material will not yield usable test data. The sizes of the holes were inspected in a microscope equipped with a crosshair reticule and a micrometer mechanical stage, and each piece was inspected after measuring the thickness of the sample over the hole area with a standard micrometer. If the use of the friction clutch in the micrometer producing a 1-pound spindle pressure on a 1/4-inch-diameter area caused any change in the measured diameter of the hole the sample was rejected. The method developed for producing round clean holes 0.003 inch in diameter to a tolerance of +0.0005 inch −0.0000 inch in Mylar insulation up to 0.008-inch thick is illustrated in Fig. 2. A balance-wheel staff from a lady's wrist watch was used for the punch. Working under a stereoscopic microscope the excess material was removed with a razor blade with the punch in place.

It was found that consistent results could not be obtained unless care was used to avoid contamination of insulation surfaces. A fingerprint on the insulation could alter results beyond use. Since the dielectric breakdown voltage of the air gap in the holes in the insulation was under investigation, temperature and humidity changes had little effect on the results. To check this point, similar voltage breakdown tests of 0.003-inch holes in 0.003-inch Mylar tape were made at 125 C (degrees centigrade) and at room temperature. The difference between the two series of tests was all within the ±20% repeatability range of the test procedure. To those not familiar with this type of test, a 20% repeatability of air-gap breakdown voltage data is not easy to achieve.

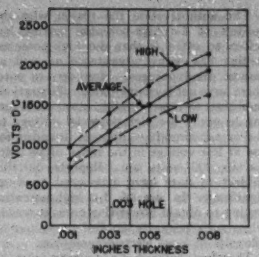


Fig. 4. D-c high-potential breakdown voltage versus insulation thickness through 0.003-inch-diameter holes in Mylar. Ten tests made at each indicated thickness; see Fig. 3 for definition of curves

Test voltages were applied to the samples with the rate controlled to agree with ASTM *D149-55T* procedure of approximately 500 volts per second.

Fig. 3 illustrates the 60-cycle a-c rms high-potential breakdown voltage for 0.003 inch diameter holes in Mylar tape 0.001 to 0.008 inch thick. Ten tests were made for each indicated thickness of tape. The solid line records the average breakdown voltage and the dashed lines show the maximum and minimum departure from this average. Fig. 4 illustrates the d-c high-potential breakdown voltage, for similar conditions, through 0.003-inch-diameter holes in Mylar tape insulation of similar thicknesses from 0.001 to 0.008 inch. Again, ten tests were made at each indicated thickness and the spread of the results from the average value shown by the solid line is indicated by the dashed lines.

A comparison of the values of d-c/a-c breakdown voltages for constant-diameter holes in varying insulation thicknesses shown in Figs. 3 and 4 would indicate that a d-c high-potential test voltage of 1.414 times a 60-cycle rms sine-wave a-c test voltage would be equivalent or, as might be expected, a d-c high-potential test voltage equal to the peak voltage of the specified rms a-c test voltage will be equivalent as far as air-gap breakdown is concerned in small-dimension insulation flaws.

Table I. D-C and A-C RMS Voltage Breakdown of Solid Mylar*

Sample No.	Test Voltage D-C	Test Voltage A-C RMS
1	3,400	1,900
2	3,300	1,950
3	3,200	1,975
4	3,000	1,950
5	2,600	2,150
6	3,100	2,150
7	3,350	1,900
8	3,150	2,175
9	3,300	2,150
10	3,250	2,100
Average	3,165	2,040

Average d-c/a-c = 3,165/2,040 = 1.55

* Insulation thickness of 0.00025 inch.

It is noteworthy that the small-hole a-c breakdown test data recorded in Fig. 3 are records of 1958 tests which are a repetition of similar tests conducted by the authors in 1956.[8] The 1958 data were obtained from tests made by different personnel and are a repetition of the 1956 tests made to ensure the validity of the a-c/d-c comparison of breakdown voltage in the present paper. It is interesting to note that the 1958 data agree with the 1956 tests with a maximum variation of the average breakdown curve of less than 5%.

High-potential dielectric tests must not only detect the small-hole air-gap type of insulation flaw but must also detect or break down marginally thin insulation. The following tests were made to provide additional data on the ratio of d-c/a-c high-potential test voltage necessary to produce equivalent results in testing for too-thin insulation. As previously mentioned, considerable investigation of solid insulation in this field has already been done.[1-5]

Mylar tape insulation of 0.00025-inch thickness was used with the same test jig and the same procedure. Table I shows the results. An examination of this table indicates that a d-c high-potential test voltage approximately 1.6 times a 60-cycle rms sine-wave a-c test voltage would be equivalent for this type of insulation and for the thicknesses mentioned. Actually, to be marginally thin, Mylar insulation for detection by the present aircraft high-potential test voltage of 1,500-volt 60-cycle rms would have to be much thinner than the minimum thicknesses available, possibly as thin as 0.0001 inch.

A d-c/a-c rms test-voltage ratio of 1.6 agrees with data obtained by other investigators. Pending further investigation of thinner insulation sections, which break down closer to the present 1,500-volt aircraft high-potential test voltage, this 1.6 ratio is recommended.[9,10]

References

1. Practical Hi-Pot Testing, **John T. Keefe.** *Electrical Manufacturing*, New York, N. Y., Aug. 1956.

2. D-C Overpotential Testing Experience on High-Voltage Generators, **R. J. Alke.** *AIEE Transactions* (*Electrical Engineering*), vol. 71, pt. III, Aug. 1953, pp. 807-70.

3. High Voltage D-C Testing of Generator Insulation. *Report*, Prime Movers Committee, Edison Electric Institute, New York, N. Y., Feb. 8, 1953.

4. Diagnoses of A-C Generator Insulation Condition by Nondestructive Tests, **A. W. W. Cameron.** *AIEE Transactions* (*Electrical Engineering*), vol. 71, pt. III, Jan. 1952, pp. 263-74.

5. Studies of Impulse Strength and Testing Problems on High-Voltage Generators, **G. L. Moses, R. J. Alke.** *Ibid.*, vol. 72, pt. III, Apr. 1953, pp. 123-31.

6. Peak Voltages With D-C Arc Interruption for Aircraft, **Virgel E. Phillips, Walter P. Mitchel.** *AIEE Transactions*, vol. 63, Dec. 1944, pp. 944-49.

7. Report of the Advisory Staff for Aircraft Electrical Systems and Appendices. Wright Air Development Center, Dayton, Ohio; Bureau of Aeronautics, Washington, D. C., available from Armed Services Technical Information Agency, Washington, D. C., Apr. 1951.

8. High-Potential Testing of Aircraft Equipment Electric Insulation, **L. B. Kilman, J. P. Dallas.** *AIEE Transactions* (*Electrical Engineering*), vol. 75, pt. II, Sept. 1956, pp. 189-93.

9. Critical Failures in Aircraft Equipment Electrical Insulation, **L. B. Kilman, J. P. Dallas.** *Electrical Manufacturing*, Mar. 1956, pp. 124-30.

10. Test Code for Aircraft Equipment Electrical Insulation. *AIEE Publication no. 803*, July 1957.

11. *ASTM Standard D149-55T*, American Society for Testing Materials, Philadelphia, Pa., pt. 6, 1955.

Discussion

J. W. Allen (Bendix Aviation Corporation, Teterboro, N. J.): The authors must be commended on their interest and efforts to correlate alternating versus direct voltages for dielectric-strength testing.

I noted the suggested ratio of 1.6 (d-c/a-c) for detection of marginally thin insulation, rather than the accepted value of 1.41. Knowing that a great deal of work had been done in determining the breakdown voltage of air, I referred to Fowle's text[1] in which paragraph 4-778 referenced a paper by Dubsky.[2] The latter contained test data on air films down to 1 mil. Reducing both sets of data to volts per mil and plotting on log-log paper, versus thickness, made a comparison easier and showed a good agreement as to slope, though differing in values, with the Kilman-Dallas data (low value of Fig. 3) approximately 30% higher than that of Dubsky.

Table II lists extracted values. Dubsky used 8- by 15-inch glass plates for film boundary. Kilman and Dallas used electrodes of approximately 0.190 inch. Tests between spheres give an indication that, for a given spacing, small spheres break down at a higher voltage than large spheres. The 0.003-inch hole used by Kilman and Dallas could be considered a very small electrode. It would be interesting to have data for a given spacing with holes of varying diameters. Breakdown values of pin holes or air films are of great significance for aircraft electric devices, since minimum weight is wanted. The data presented indicate that multilayer insulating films would increase the reliability of insulation.

Table II. Data Comparison

Source	Temperature	Volts Per Mil 1 Mil	2 Mils	4 Mils	6 Mils	8 Mils
Kilman and Dallas (from Fig. 3)	high	765	500	320	242	197
	low	470	330	230	181	151
Dubsky	25 C	340	253	180	144	123
	100 C	232	205	148	120	100

The inference in the paper that voltage breakdown at 25 C and 125 C cannot be separated because of a repeatability range of ±20% does not agree with 500 successive tests made by Hayden and Eddy.[3] It is stated that the expected maximum error of six tests is ±2.9%.

It is noticeable that many present-day papers do not list references of "mellowed vintage." The oldest reference cited is of the year 1951. To extend the reference area, for the benefit of those interested in the subject, references 4 through 8 are given. These also contain references to additional information.

If my remarks can stimulate the present (or future) authors to hunt for and delve into the information available, I am certain the added background will aid them.

REFERENCES

1. STANDARD HANDBOOK FOR ELECTRICAL ENGINEERS, F. F. Fowle. McGraw-Hill Book Company, Inc., New York, N. Y., sixth edition, 1933.

2. THE DIELECTRIC STRENGTH OF AIR FILMS ENTRAPPED IN SOLID INSULATION, F. Dubsky. *AIEE Transactions*, vol. 38, pt. I, 1919, pp. 537–75.

3. THREE THOUSAND TESTS ON THE DIELECTRIC STRENGTH OF OIL, J. L. R. Hayden, W. N. Eddy. *Ibid.*, vol. 41, June 1922, pp. 394–401.

4. LAWS OF ELECTRICAL DISCHARGE IN GASES, W. R. Carr. *Philosophical Transactions*, Royal Society of London, London, England, Series A, vol. 201, 1903.

5. CONDUCTION OF ELECTRICITY THROUGH GASES (book), V. J. Thompson. Cambridge University Press, Cambridge, England, 1906.

6. ELECTRICITY IN GASES (book), J. S. Townsend. Oxford University Press, London, England, 1915.

7. COMPARISON OF AC AND DC SPARK POTENTIALS, J. C. Jansen. *Physical Review*, New York, N. Y., vol. 8, 1916.

8. IONIZATION OF OCCLUDED GASES IN HIGH-TENSION INSULATION, G. B. Shanklin, J. J. Matson. *AIEE Transactions*, vol. 38, pt. I, 1919, pp. 489–536.

W. B. Penn (General Electric Company, Erie, Pa.): The authors are making a worthy contribution in their attempts to resolve the provoking problem of comparing a-c high-potential testing with d-c testing. The question to be resolved, however, is whether the overly simplified system of sheet Mylar between two electrodes, evaluated by the authors, is at all representative of or comparable to the system in actual use in electric rotating machinery. These systems consist of composite mica slot insulators, enameled or fiber wire insulators with an over-all resin impregnation, and are systems wherein the dielectric constant or specific inductive capacity of each of the materials involved is different. Therefore, a voltage imposed across these systems presents a complex nonuniform voltage stress distribution across this high-resistance path of insulations in series not at all comparable to the perforated Mylar test system.[1] A comparison between a-c and d-c breakdown on a composite insulation would be much more valid and convincing. However, it would be surprising if the 1.41 d-c/a-c ratio established on the Mylar were also found on the composite system.

Corona-start voltage should be a more sensitive tool for detecting when imposed voltages are approaching close to the breakdown point than rapid increase of leakage current, since leakage at 500 volts or 1,000 volts cannot normally be measured with the conventional microammeter on insulation of good quality, and voltages much above these values become unrealistic for testing equipment designed for operation at 30 volts. For high-voltage machines, or machines in poor condition, or damp, the leakage-current procedure does offer some promise as an exploratory tool.

We agree that a-c high-potential tests will result in different values than d-c high-potential tests. However, the authors have failed to show that d-c tests are any more reliable than a-c, which have at least a background of usage, or that d-c will locate incipient faults which a-c might overlook. Further investigation on systems more nearly approaching those in actual use should be more indicative of the value of d-c high-potential testing.

REFERENCE

1. PREDICTING INSULATION FAILURES WITH DIRECT VOLTAGE, C. M. Foust, B. V. Bhimani. *AIEE Transactions*, vol. 76, pt. III, Dec. 1957, pp. 1120–39.

R. L. Balke (General Electric Company, Erie, Pa.): The authors indicate that the prime reason they selected thin Mylar sheet insulation for their tests was to obtain comparative data on air-gap type of faults and marginally thin insulations that could be detected at voltage levels near 1,500 volts. However, the selected material and manner in which it was tested is not representative of the types of insulations normally employed in the majority of aircraft equipment. For instance, windings of rotating machines are normally insulated from ground by a rather complex system comprised of air, composite 2- and 3-layer mica-type insulators, enamels or fibers on the windings, and an over-all impregnating resin. Data presented by Foust and Bhimani (see reference 1 of Mr. Penn's discussion) indicate a significant difference exists in the distribution and ionization effect of direct versus alternating voltages applied to such complex systems. Would the authors comment on the applicability of these type data to the more complex aircraft insulations and, if believed applicable, how they would propose to apply them.

The authors state the reason why the subject of d-c high-potential testing has arisen so often is that d-c high-potential testing has many advantages as well as serious disadvantages. However, to date, there have not been any data or illustrations published to present valid reason for wanting to apply d-c high-potential voltage to d-c or a-c aircraft equipment intended for use at low voltage in the range of 28 volts d-c to 200 volts a-c. Reports on this subject to date have attempted to show significance in use of d-c high-potential testing to obtain absorption current trends when such tests are applied at values of kilovolts to large electric apparatus.

Therefore, before a d-c/a-c ratio is established, it would seem more justification should be demonstrated to warrant use of d-c high-potential testing on low-voltage aircraft equipment. Should such testing show significant advantage, perhaps there may be more reason to discontinue a-c testing rather than to establish ratios. Similarly, if the d-c testing does not have virtue for low-voltage equipment, it would seem inappropriate to apply it. The authors opinion, as to the validity of d-c high-potential testing of low-voltage aircraft equipment, might focus more attention to resolution of this controversial subject.

G. L. Hill (Hill Research Company, Oakland, Calif.): The authors have presented some interesting data on relatively low direct voltage tests on insulation for aircraft equipment. The writer's original work[1,2] was with the maintenance testing of large rotating electric machinery of 2,300-volt a-c rating and above. It was definitely determined from recorded test data covering a period of about 10 years' test experience on 4 million kva of apparatus that the approach to the breakdown point in insulation could be predicted. Subsequently it has been determined by the writer and others that the same characteristic test results may be obtained on low-voltage apparatus such as 120/240/440-volt a-c motors, d-c generators, and d-c motors.

The writer, although not especially familiar with aircraft equipment, cannot see any reason why similar characteristics should not be obtained. Fig. 5 shows the test data and curve obtained on a 1-horse-

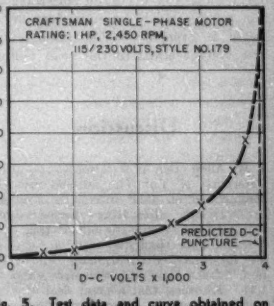


Fig. 5. Test data and curve obtained on single-phase 120/240-volt a-c motor

power 120/240-volt 60-cycle motor. This motor had about 5 years of intermittent shop duty before testing and has had about 2 years since the test. It will be seen that the test was stopped at 3,700 volts d-c and the predicted puncture value would be somewhere around 3,900 or 4,000 volts d-c. The insulation was not damaged as it has been operating on 220 volts a-c since the test.

The authors elected to use a small test jig, shown in Fig. 1, for testing to obtain the breakdown voltage of thin sheet insulation with artificially produced faults. They point out that punching holes in thin insulation for test purposes is a real problem. The writer concurs fully with this and considers there is no known artificial method of punching holes in thin insulation that will actually simulate deterioration of electrical insulation in service. Therefore erroneous conclusions may be drawn by artificially damaging insulation and making tests thereon. The writer would like to see the authors repeat their work on complete aircraft equipment that has been in service for varying periods and duty cycles.

A very important value of d-c nondestructive testing is on maintenance work where it is desirable to determine the dielectric condition of the insulation, without puncturing. It would seem that it should be a valuable tool for that purpose when the importance of reliable operation of all of the component parts of the electric equipment on an airplane is considered.

If the authors desire to test new sheet insulation, it is recommended that a much larger area be used, such as sheets with a layer of conducting material painted or pressed on each side. A guard ring should be used to avoid errors due to surface leakage currents. The currents obtained for varying applied voltages are not given in the paper but they must have been in the order of micromicroamperes for a small 1/4-inch electrode. If the currents are this small the most minute contamination, humidity, etc., will cause a greater variation in current than the conduction current passing through the insulation.

It is impossible to obtain a wide contact area with small metal rod electrodes, therefore it is possible that an infinitesimal point contact would overheat this path, cause melting due to the low thermal conductivity of the material, resulting in puncture without warning.

References

1. Testing Electrical Insulation of Rotating Machinery with High Voltage Direct Current, G. Leslie Hill. *AIEE Transactions*, vol. 72, pt. III, Apr. 1953, pp. 159–74.

2. Method of Testing Electrical Insulation. *U. S. Patent no. 2,837,714.* G. Leslie Hill.

L. B. Kilman and J. P. Dallas: The authors agree with Mr. Allen that the breakdown voltage of various open air gaps has been well established, but they disagree that these data are applicable to the conditions that exist in common insulation faults such as small-dimension holes or cracks. Voltage stress distribution is always the key to breakdown voltage values. The change in dielectric constant from insulating material to air at the edges of a 0.003-inch-diameter hole or crack materially affects the voltage distribution in the air gap within the hole and therefore the resultant breakdown voltage. As insulation flaws decrease below 0.003-inch diameter, the breakdown voltage through the flaw will increase rapidly, approaching a value close to the voltage breakdown value of the solid insulation for flaws with a major diameter of less than 0.001 inch. Thus the breakdown voltage of 0.001-inch flaws may exceed 1,000 volts per mil. or more than twice that of free air gaps. For holes larger than 0.003-inch diameter the breakdown voltage approaches the breakdown value of similar open air gaps at around 0.010-inch diameter. This "small-hole effect" on air-gap breakdown voltage through holes in insulation is considered the key factor in determining high-potential test voltages.

For the last 15 years the authors have directed the operation of an aircraft laboratory one of whose main functions has been to analyze the cause of failures occurring in aircraft electric equipment in use. Small-diameter insulation flaws of the order of 0.003 inch have been the major problem, especially where high-potential test voltages below the values necessary to detect such flaws are specified and used.

The authors wish to thank Mr. Allen for bringing out the point that small-hole breakdown voltage will vary with temperature in the conventional manner. This point is not made sufficiently clear in the paper. The intention was to state that this factor is small compared with the other factors affecting the observations.

The references studied in connection with the various papers the authors have written on this subject would form a bibliography exceeding by many times the space required to publish this paper. Our earliest reference is to an 1830 section of Michael Faraday's diary. In compiling the list of references for this paper the authors chose to list only those references which bore directly and most significantly on the material under discussion. Many of the questions raised in the comments on this paper, including Mr. Allen's request for more data on the effect of hole diameter on breakdown voltage, are discussed in the authors' previous papers on this subject which are listed in references 8 through 10 of the paper.

With reference to Mr. Penn's, Mr. Balke's, and Mr. Hill's comments: electric discharge through small-hole air-gap flaws in insulation is not materially affected by the type of insulation for common materials with dielectric constants in the range of two to five. Therefore a comparison of a-c/d-c breakdown voltages for small-hole faults in composite insulation, suggested by Mr. Penn, is not expected to yield results that differ substantially from those shown in this paper. In considering the breakdown of marginally thin insulation the effects of composite insulation with varying dielectric constants can be substantial and so varied and complex as to defy experimental determination. It should be pointed out that the 1.41 d-c/a-c breakdown ratio was established only for the air-gap breakdown condition through small-hole insulation faults and that varying higher ratios were found to exist for all solid insulation faults which have been investigated, either by the present writers or the references.

Mr. Penn's suggestion of corona-start voltage as a more sensitive tool for detecting insulation faults is an impractical approach to the problem of sea level testing of aircraft equipment insulation which is the exclusive subject of this paper. At voltages under 2,000 volts, which appears to be the very maximum that can be recommended for aircraft equipment tests, corona conditions will not be approached for flaws in common thicknesses of aircraft equipment insulation.

It is to be noted with reference to both Mr. Penn's and Mr. Balke's comments, that the use of d-c high-potential voltages is not being "justified by the authors." We are attempting to answer the questions already raised by the industry as to the direct voltage which would be equivalent to specified a-c test voltages. With electric machines of 60 kilowatt or larger, the advantages of d-c high-potential testing may outweigh its disadvantages.

Referring to Mr. Hill's comment, it may be noted that AIEE Publication no. *803* (see reference 10 of the paper) states: "It may be further noted that many of the better aircraft insulations now in use do not show any significant increase in leakage current prior to rupture. While nondestructive high-potential testing is desirable, there is not at present sufficient evidence to warrant describing any present d-c or a-c equipment or method as nondestructive."

The principal purpose of high-potential testing is to detect insulation flaws and marginally thin insulation. In ordinarily used thicknesses and types of insulation, breakdown through these flaws occurs long before the rise in general leakage current mentioned by Mr. Hill. It is never the intention in testing aircraft equipment to test the general insulation to destruction or to indicate the point at which the healthy insulation would fail as proposed by Mr. Hill. We do need to detect the flaws in unhealthy insulation which do not always show a significant rise in d-c leakage current prior to breakdown. The important point which Mr. Hill and other advocates of so-called d-c nondestructive testing miss is that the rise in leakage current near the general breakdown of insulation is a large-area phenomenon whereas insulation failure in aircraft equipment is generally due to faults involving very small areas which in general break down long before the main area of insulation is stressed sufficiently to produce a distinguishable rise in leakage current. To make this point very plain let us repeat: 1. Modern insulation such as teflon, Mylar, and many of the silicon compounds do not show a measurable increase in leakage current as the breakdown point is approached for the total insulation areas involved in aircraft equipment. 2. Even with insulations which do show this effect for the areas involved in aircraft equipment the prebreakdown rise in leakage current is only a function of large-area near-breakdown stressing of the insulation and is not observable where a fault exists in the insulation involving the dielectric tressing of a relatively small fault area.

Development of a Dimmable Aircraft Fluorescent Light System

E. W. BANIOS
ASSOCIATE MEMBER AIEE

THE DEVELOPMENT program to be described was for the Douglas *DC-8* commercial jet airplane and was inaugurated to provide an effective dimmable aircraft fluorescent light system. Prior to this development none was available. The deficiency had made fluorescent lighting for aircraft unattractive, despite its many other advantages. Specifically, the purpose of this program was to design and develop two dimmable systems: one for the cabin which would have a maximum lamp current of 150-ma and an extremely high dimming ratio, and one for the cockpit which would have a maximum lamp current of 200-ma (very high light level) and a moderate dimming ratio.

Design Problems

Confronting the program were several difficult design problems, which apparently were largely responsible for preventing the development of any systems. These problems are described in the following.

The *T-5* type of fluorescent lamp yields the highest light output to weight ratio of all other medium-sized lamps. For this reason it was chosen for use in the airplane, but along with this decided advantage the unstable electrical characteristics common to all forms of gaseous electric discharge lamps are present to a highly exaggerated degree. These result in a tendency of the lamp light to fluctuate in a most objectionable manner when attempts are made to dim it.

Three general phenomena are responsible for these unstable characteristics:

Jitter

Jitter is a random and abrupt variation of the lamp light caused by inconsistent lamp reignition each semicycle. It is due to the shallow slope of the lamp voltage as it crosses the lamp ignition voltage range, and the random variation of the actual ignition voltage value in this range.[1] This variation of the lamp ignition voltage is caused by the variation in residual lamp gas ionization, which in turn is dependent on the lamp gas temperature and an arc plasma factor that has been suggested by experimental observations. The factor is proportional to the arc cross-sectional area and inversely proportional to the arc plasma length and circumference. This mathematically reduces to a ratio of plasma diameter to length. The lamps which have large plasma factors exhibit less tendency to jitter.

Since the *T-5* lamp has a relatively small plasma factor and mass (thermal inertia), significant variations in residual ionization occur, and unless the lamp voltage traverses this region rapidly (steep wavefront slope) random variations in the phase of lamp reignition results and jitter is manifested. Lamp gas pressure also has some effect, but it is minor in comparison.[1]

Oscillations

Voltage oscillations across the lamp, and thus current and light output oscillations, are mainly caused by space charge energy oscillations occurring during transition from glow to arc lamp operation. Fig. 1 illustrates the glow and arc regions on the lamp characteristic. It is obvious from this that the lamp must pass through glow to arc operation during all but extremely low current flow operations. The energy which was in the glow prior to arc operation commencement must be dissipated, and oscillations take place until an equilibrium is reached. These oscillations, however, are in most part regular and thus not observable if the reignition phase is consistent.[2]

Striations

Striations are light and dark bands in the arc plasma apparently moving longitudinally at various speeds. One explanation of the phenomenon deals with the Druyvesteyn and Penning cumulative ionization theory: Ionization is facilitated by the concentration of the plasma in sections, each containing regions of high and low field strength. Since it has been found that the potential across each of these striations is less than the ionizing potential of the gas, Arnott theorizes that metastable ions are being formed ahead of each ion bunch, thus causing potential concentration in the ionized (or bright) region and a propagation toward the cathode. The effect is less at higher currents and eventually a current value is reached where a homogeneous plasma is more efficient.[3]

The striation velocity is in the order of ten times the ionic drift velocity, and therefore not observable. It is the variations in the striation velocity that make their effect on the lamp light output visible. The causes of these variations are the same as those for lamp ignition volt-

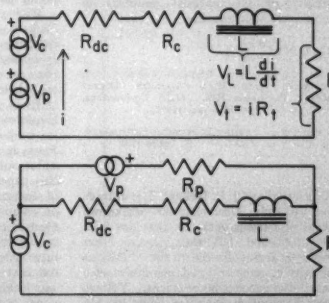


Fig. 1. Approximate equivalent circuits for basic fluorescent lamp-dimming system

Paper 58-839, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted February 13, 1958; made available for printing May 19, 1958.

E. W. Banios, formerly with the Douglas Aircraft Company, Santa Monica, Calif., is with the Space Technology Laboratories, Inc., Los Angeles, Calif.

The author expresses his appreciation for the cooperation and assistance the Engineering Departments of the Mechapoments Division of Servomechanisms, Inc., Sylvania Electric Products, Inc., and the Miniature Lamp Department of General Electric Company exhibited in this development project. Appreciation and thanks are also extended to Charles J. Jenkins, mechanical design specialist the Douglas Aircraft Company, Inc., for his counsel, assistance, and encouragement during all phases of the project. The work performed by Arthur Bastian of the Ward Leonard Electric Company, described in reference 5, was the foundation of this development.

Fig. 2. Thirteen-watt T-5 type of fluorescent lamps: top lamp is standard bi-pin type and bottom two lamps are ring-and-button, rapid-start type

age variation, and since the *T-5* lamp has small thermal inertia and plasma factor it exhibits the effects of striations. These striations were found not to be objectionable if their variation is regular or harmonic. It is when they vary in a random manner that the lamp light output is adversely affected. Several unsuccessful attempts were made by the lamp manufacturers to minimize these unstable characteristics by minor lamp redesign, and the problem was then relegated to solution by system design.

Another problem involving the lamps arose when tests indicated that the commercially available *T-5* type of lamp could not meet the airplane environmental vibration requirements, failure occurring at the electrode filaments. The lamp manufacturers ably met the situation by developing ruggedized filaments which satisfied the vibration requirements. These lines of aircraft *T-5* fluorescent lamps have now been made available commercially; see Fig. 2.

In addition, there was the ever-present problem of weight and size minimization. Adequate reliability and realistic service life had to be balanced against the system weight.

Design Attack

The basic rapid-start fluorescent lighting principle was used because it lent itself to dimming application, and eliminated the starter. The lamp filaments are continuously energized during this type of operation, making available by thermionic emission an abundance of electrons for the lamp arc. By these means the voltage necessary to arc the lamp is reduced by approximately 50%.

The basic requirements for a successful rapid-start dimming system are as follows:

1. The filaments must be continuously energized to maintain hot cathode operation.

2. The lamp must be reignited each semicycle by a voltage of sufficient amplitude impressed across it.

3. The lamp must be properly ballasted by a series impedance during the lamp arcing period (due to the negative slope of the lamp resistance characteristic).

4. The energy delivered to the lamp by the current flow each semicycle must be controllable over a wide range (dimming).

5. The lamp light output over the entire dimming range must not be physiologically, aesthetically, or psychologically objectionable.

These requirements were satisfied as follows (in the same order):

1. A transformer with isolated secondary windings easily satisfied this requirement.

2. A saturating type of peaking transformer was used here for reasons of weight and reliability.

3. An inductive reactor was used to satisfy the ballasting requirement because it lent itself to the consolidation of the auxiliary unit which is described later.

4. The light level was controlled by varying and controlling the voltage magnitude across the lamp-ballast series circuit.

5. The quality of the light output is in reality the criterion for system success. Measures to meet this requirement dealt exclusively with the lamp instabilities previously discussed and defined. These measures were:

a. A steep lamp voltage wavefront of sufficient amplitude to reignite the lamp at the same phase angle each semicycle was provided to eliminate jitter (peaking transformer design).

b. High-series ballasting impedance at low and medium lamp current operation was effected to minimize the striations (peaking transformer secondary-ballast design).

c. The proper lamp voltage waveshape was provided to insure that no discontinuities existed throughout the dimming range-smooth transition of lamp voltage waveshape as lamp current increases or decreases (for over-all design, see the section "Lamp Voltage Waveshapes)."

All information pertaining to the lamp instabilities indicate that the striation anomaly cannot be overcome completely.[1,2] The foregoing measures have minimized them to an acceptable degree.

System Description

Fig. 3 depicts the schematic arrangement of the basic system. Fig. 4 specifically defines the circuit for the basic system. This circuit combines the peaking and filament transformer primary windings, and the peaking transformer secondary and ballast coil windings. This arrangement permitted consolidating the auxiliary unit as shown in Fig. 5.

The purpose of the ground plane in close proximity to the lamp indicated in Fig. 4 is to lower the lamp ignition voltage by increasing the electrical stresses in the gas in the vicinity of the high potential electrode.

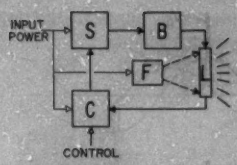


Fig. 3. Block diagram of fluorescent lamp-dimming system concept

S = saturating transformer
B = series ballast
F = filament transformer with isolated secondary windings
C = control transformer with variable output voltage
L = fluorescent lamp

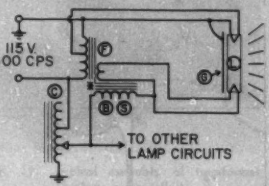


Fig. 4. Basic circuit for fluorescent lamp-dimming system. See Fig. 3 for symbol explanation

* Saturable magnetic linkage

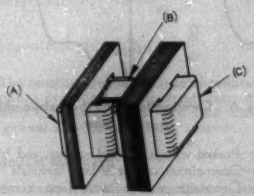


Fig. 5. Physical arrangement of transformer-ballast unit

A—Peaking transformer secondary and ballast coil
B—Saturating core
C—Filament transformer and peaking transformer primary coil

Lamp Voltage Waveshapes

The importance of establishing the proper lamp voltage waveshape in the circuit has been briefly mentioned. To the author's knowledge, there is no pub-

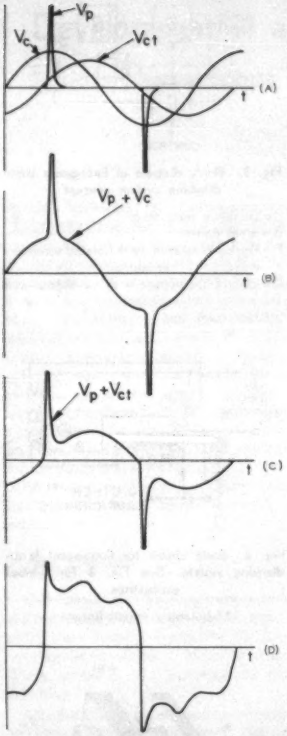


Fig. 6. Circuit voltages and their relationship to each other

A—Peaked voltage, control voltage, and V_{ct}
B—Open-circuit voltage at lamp terminals
C—Composite voltage with lamp current flowing: V_p+V_{ct}
D—Characteristic fluorescent lamp voltage waveform for a-c operation

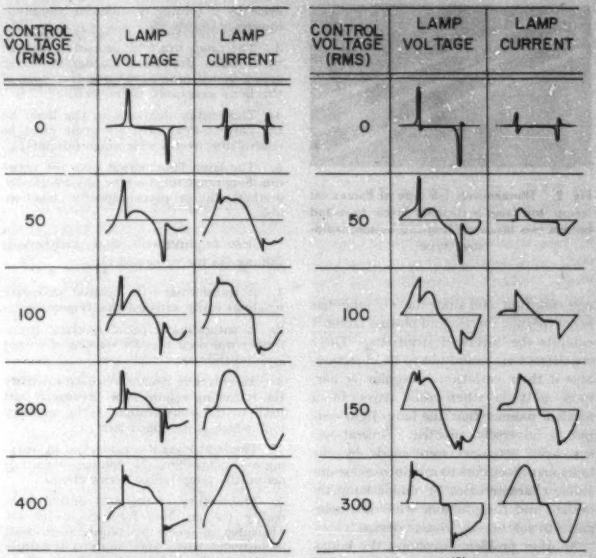


Fig. 7. Lamp voltage and current waveforms of the 150- and 200-ma maximum lamp current version of the 13-watt T-5 fluorescent lamp dimming system. Waveforms are as observed on an oscilloscope and, due to extreme variation of lamp current amplitudes over dimming range, no consistent scale is used

A—150 ma maximum B—200 ma maximum

lished information regarding this subject. It was found that its consideration in fluorescent lamp dimming system design is very important, and for this reason the following discussion pertaining to lamp voltage waveshapes is included.

Fig. 6 shows the development of the composite driving voltage (peaking plus transferred control voltage) which is impressed across the lamp. In the course of experiment with the design, it was discovered that the lamp instabilities are reduced when the composite voltage closely approximates the lamp voltage waveshape at medium lamp current, particularly the coincidence of the peaks. This waveshape, displayed in Fig. 6(D), is typical for fluorescent lamps with series impedances operated with an alternating voltage source.[2] The Appendix discusses this matter further.

Fig. 7 displays the various waveforms of the two system designs: Fig. 7(A) shows the 150-ma (milliampere) maximum lamp current version for cabin lighting and (B) shows the 200-ma maximum lamp current version for cockpit lighting. The ballast coil of the 150-ma unit could be wound with more turns than the 200-ma unit because it was possible to use smaller wire gage. Less saturating core material was thus required to develop the same peaked voltage amplitude. This caused the saturation point to occur earlier in time; see Fig. 8.

Fig. 7 shows the changes in lamp voltage waveshapes with increasing control voltage for the two versions. It is interesting to note, in the 200-ma version, the suppression of the peaked voltage, and the rise of the peak of the characteristic lamp voltage commencing at 150-volt control voltage. The transition of the lamp voltage for the 150-ma version (in which the composite voltage more closely resembles the typical lamp voltage) is smooth with no radical change to the peak. At the point of the peaked voltage suppression in the 200-ma unit, moderate striations were observed in the lamp operation.

The actual mechanism responsible for the lamp striations is not sufficiently known to permit precise quantitative design for its minimization. For this reason the requirements for the composite voltage waveshape cannot, at this time, be analytically justified; however, experimental evidence does support it.

Conclusions

Figs. 1 and 9 and the Appendix describe and discuss the system's operational dependence on the interaction of its nonlinear components. It is because of this situation that standard design techniques could not be effectively employed. A procedure embodying qualitative opera-

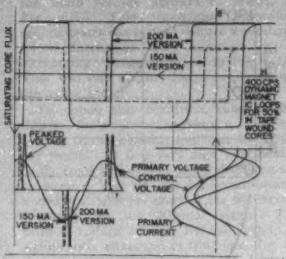


Fig. 8. Graphic analysis of peaked voltage generation

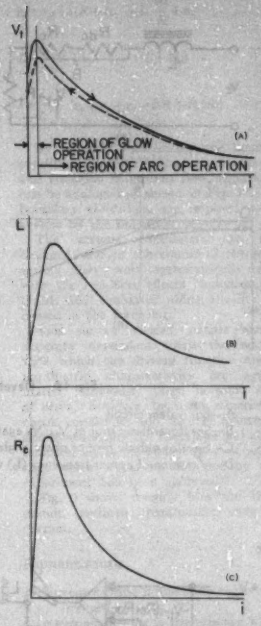


Fig. 9. Variation of instantaneous values of the basic circuit parameters with lamp current

A—Lamp voltage
B—Series ballast inductance
C—Effective core resistance: fictitious resistance assigned to core loss effect

Fig. 10. Transformer-ballast unit: top is a unit prior to impregnation and potting, and bottom is a finished unit

Fig. 11. Typical assembly of transformer ballasts as installed

tional analysis and quantitative trial and error technique was used, and results were found to be good, though tediously obtained. The associated mathematical analysis is contained in the Appendix.

Of the two system versions considered here, the 150-ma version (main-cabin lighting) possessed the greater dimming ratio and lamp stability. This version, however, was not completed because of a design decision to change to another form of cabin lighting. The 200-ma version, intended for use in the cockpit overhead and main instrument panel floodlighting scheme, was completed and is now installed and being successfully used. Its dimming ratio, far less than the 1,000 to 1, indicated for the 150-ma version, is only 100 to 1; however, this dimming ratio is found to be quite adequate for the cockpit application.

This development program culminated in the design of an 8-watt and 13-watt *T-5* fluorescent-lamp dimming auxiliaries (a 13-watt unit is shown in Fig. 10). Fig. 11 shows an assembly of six 8-watt units as they are installed on the airplane.

Investigation indicates that two other basic approaches to the problem of fluorescent lamp dimming for aircraft have good potential. One involves use of a high-frequency power supply, in the order of 4 kc. The advantages of this type of system would be the increased lamp efficiency obtained at higher frequency operation, and the decrease in size and weight of the auxiliary components. One potential serious disadvantage is the possible increase in radio noise generation. This could either make the system not at all usable in the airplane or could significantly increase the total system weight by the addition of shielding and filtering equipment.

The other possibility is a system very much like the one described in this paper, except that the control is accomplished by saturating the series impedance ballast. One obvious advantage this system might have is the decrease in weight, size, and volt-ampere capacity of the dimming control unit.

As a result of this development program, there now exists a better understanding of the fluorescent lamp instabilities associated with dimming operations, and the design steps required to minimize them. In addition, the seeds were sown for a practical design technique on a rather moderate academic level to cope effectively with nonlinear systems.

Appendix. System Analysis

Mathematical Approach

The system's approximate equivalent circuits are displayed in Fig. 1 with the circuit quantities and parameters defined in the nomenclature. The following equations are derived by summing voltages around the series circuit:

$$v_c + v_p = i[R_{dc} + R_c + R_l] + L\frac{di}{dt} \qquad (1)$$

$$v_c + v_p = i[R_{dc} + f(i) + g(i)] + h(i)\frac{di}{dt} \qquad (2)$$

The functions $f(i)$ and $g(i)$ are nonlinear variables dependent on the circuit current; therefore, equation 2 does not lend itself to general solution or obvious analysis.

Practical Approach

To obtain any kind of analysis, a less rigorous, and therefore less accurate, method of approach must be employed. The actual accuracy of this approach will depend on the closeness of the approximations and the applicability of the assumptions made, but it is apparent that sweeping assumptions and broad liberties must be made and taken if any kind of general understanding is to be obtained.

Step 1. Generation of v_p: The primary current in the peaking transformer is assumed to be sinusoidal. The justification for this is that the total core of the peaking-filament transformer need not saturate. Only the small parallel segment (the saturating core) of the total magnetic circuit saturates. The phase angle between the primary current and the voltage is assumed to be in the neighborhood of 75 degrees. This was experimentally observed and attributed to transformer leakage reactance.

Fig. 8 illustrates graphically the generation of v_p, which is dependent on the derivative of the saturating core flux with

respect to time. Also, its slope is dependent on the second derivative of the saturating core flux with respect to time. This permits a very steep wavefront, because the change of flux direction is extremely abrupt, due to the "squareness" of the core material magnetic characteristic dynamic loop. As stated previously, this steep wavefront is one of the requisites for eliminating lamp jitter and minimizing lamp striations.

During the lowest level of light output, the circuit-driving voltage is the peaked voltage alone; therefore, the energy delivered by the peaked voltage determines the dim point in the systems' operating range. If the flux were a square wave (which it does approach), the area under the curve which describes the derivative of this flux is as follows:

$$\text{area} = \mathcal{L}^{-1}\left[\mathcal{L}\int_0^{T+\epsilon}\left(\frac{dA(T)}{dt}\right)dt\right] \tag{3}$$

where $A(T)$ is a step function which goes from zero to A amplitude at time T, and ϵ is an infinitesimally small quantity.

$$\text{area} = \mathcal{L}^{-1}\left\{\frac{1}{S}\left[\int_0^{T+\epsilon}\left(\frac{dA(T)}{dt}\right)dt\right]_{t=0} + \frac{1}{S}\mathcal{L}\left(\frac{dA(T)}{dt}\right)\right\} \tag{4}$$

$$= \mathcal{L}^{-1}\left\{0+\frac{1}{S}\left[-(A(T))_{t=0}+S\left(\frac{A}{S}\right)\right]\right\} \tag{5}$$

$$= \mathcal{L}^{-1}\left(\frac{A}{S}\right) \tag{6}$$

$$= A \tag{7}$$

where S is the Laplacian operator.

The energy delivered by a voltage operating across a constant electric resistance is directly proportional to the voltage squared, integrated with respect to time, or the area under the voltage squared curve.

These two relationships can be applied to the peaked voltage, and the following conclusion thus reached: the energy delivered to the lamp during the dimmest level of operation varies with the area under the peaked voltage when plotted against time.

Less magnetic material was used in the saturating core for the 150-ma system than for the 200-ma system. The magnetic loop, flux, and peaked voltage for the 150-ma system is displayed in Fig. 8 with dashed lines. It may be seen from this why the 150-ma version could be dimmed to a lower light level: the peaked voltage has less area under it. The 150-ma system also yielded better lamp stability and performance. This appears to be due to greater series ballasting permitted by the additional turns of wire possible in the saturating secondary-ballast coil because of the use of smaller gage wire; and the advancement in time of the peaked voltage. This latter effect will be discussed later.

Step 2. Generation of v_{ct}: The voltage v_{ct} is defined as that voltage appearing across the lamp which is solely due to the control voltage v_c. Fig. 12(A) depicts the equivalent circuit which is pertinent to the generation of v_{ct}. Assuming that the parameters of the circuit remain constant for the sake of approximate analysis, the following transfer function is derived:

$$\frac{V_{ct}}{V_c} = \left(\frac{R_t}{L}\right)\frac{1}{\left(s+\frac{(R_t+R)}{L}\right)} \tag{8}$$

The gain and phase-angle curves for this function are plotted in Fig. 12(B) and (C). Physically, this means that the control voltage is attenuated (series ballasting) and that it displaced amounts dependent on the transfer function.

The corner frequency (f_0) for this transfer function is equal to $(R_t+R)/2\pi L$. The more f_0 is less than 400 cycles per second, the power frequency, the closer the phase displacement between v_c and v_{ct} approaches 90 degrees.

Step 3. Phase relationship between v_p and v_{ct}: Considering the foregoing, the phase relationships between v_p, v_c and v_{ct} are displayed in Fig. 6(A). The open-circuit voltage across the lamp terminals is shown in Fig. 6(B). With no current flowing in the circuit, no phase displacement occurs between v_c and v_{ct}. As soon as some finite R_t is effected in the circuit,

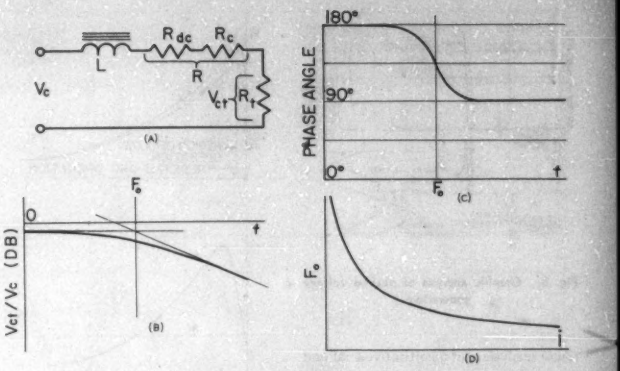


Fig. 12. Development of V_{ct}

A—Equivalent circuit
B—Semilogarithmic plot of V_{ct}/V_c against frequency
C—Semilogarithmic plot of phase angle between V_{ct} and V_c against frequency
D—Variation of corner frequency (f_0) with lamp current

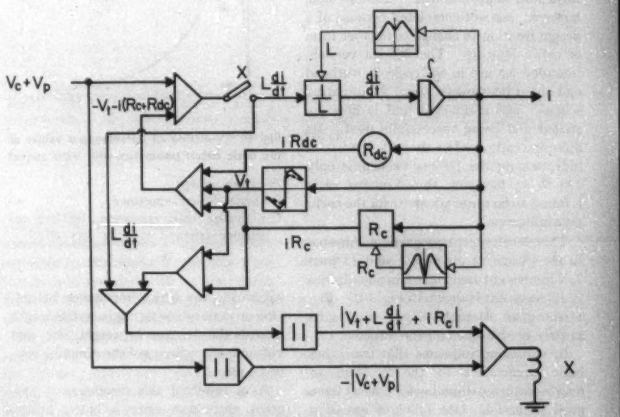


Fig. 13. Analog diagram of basic fluorescent lamp-dimming system. X picks up on positive signal and drops out on negative signal

the lamp voltage appears as shown in Fig. 6(C). This waveshape somewhat approximates the typical voltage waveshape for a fluorescent lamp operated by an alternating voltage as shown in Fig. 6(D). This was one of the design requirements discussed earlier.

To obtain this condition, close to 90-degree displacement between v_c and v_{cl}, f_0 must be considerably less than 400 cycles per second. If L is very high, comparable to R_l, some appreciable phase displacement between v_c and v_{cl} can be accomplished; however it falls far short of the 90 degrees necessary for the 200-ma system. The peaked voltage in the 150-ma system is significantly advanced in time (see Fig. 8), and therefore less displacement between v_c and v_{cl} is required to effect the desired lamp voltage waveshape. This condition also appears to explain the superior performance of the 150-ma system.

Analog Survey

Although satisfactory operating designs have been obtained without the aid of rigorous mathematical analysis, the dependence on experimental data makes final evaluation of the design awkward and extremely time-consuming. It is essential to determine exactly how and to what degree the variations in the circuit parameters which can be expected in production will affect the system operation. Also, it is highly desirable to determine if the design is in fact the optimum, or if combinations of changes in the circuit parameters would improve the over-all design. Fortunately, these conditions can be checked by survey on the analog computer.

By rewriting equation 2, the following is obtained:

$$v_c + v_p = i(R + R_c) + L\frac{di}{dt} + v_l \qquad (9)$$

$$L\frac{di}{dt} = v_c + v_p - v_l - i(R + R_c) \qquad (10)$$

$$i = \frac{i}{L}\int_0^t [v_c + v_p - v_l - i(R + R_c)dt] \qquad (11)$$

By properly establishing the boundary conditions for this integral, commencement and cessation of current flow, equation 11 can be analoged as shown in Fig. 13. The boundary conditions are imposed on the system by the operation of relay X.

The various parameters are easily changed, and in the course of the analog survey they were systematically varied with the resulting effects compared, analyzed, and evaluated along the lines discussed in the foregoing.

This survey yielded results with an accuracy dependent upon the accuracy with which the driving function and the parametric characteristics are approximated. Obviously there is some degree of error; however, by using experimental check conditions and care in determining and synthesizing the functions and characteristics, an order of accuracy of at least that which would be obtained by model experiment has been achieved.

Fig. 9 shows roughly how the 13-watt circuit nonlinear parameters vary with current.

Nomenclature

i = current flowing in the lamp-ballast series circuit

i_{pr} = current flowing in the primary winding of the peaking-filament transformer

v_{pr} = voltage across the primary winding of the peaking-filament transformer

v_p = peaked voltage generated across the peaking transformer secondary-ballast coil winding

v_c = control voltage applied to the lamp-ballast series circuit, and in phase with v_{pr}

v_{cl} = that portion of the voltage impressed across the lamp which is solely due to v_c

V_c = Laplace transform of v_c

V_{cl} = Laplace transform of v_{cl}

R_{dc} = d-c resistance of the operating circuit (considered constant)

R_c = resistance equivalent to one which represents the core losses of the inductor ballast (a function of i)

R_l = resistance of the fluorescent lamp (a function of i)

R = the sum of R_{dc} and R_c

L = the inductance of the ballast component (a function of i)

f = frequency in cycles per second

References

1. Flicker of Sign-Tubing "Halolight" Lamps. J. F. Waymouth. *Report no. E-663*, Sylvania Engineering, New York, N. Y., June 23, 1953.

2. Noise in Gaseous Discharge Lamps. J. Warren Culp. *Illuminating Engineering*, New York, N. Y., Jan. 1952.

3. *Report 221*, J. Warren Culp, William Hoenig. Project no. *113–4136*, Sylvania Physics Laboratory, New York, N. Y., Aug. 1, 1950.

4. Dimming Hot Cathode Fluorescent Lamps. J. H. Campbell, H. E. Schultz, W. H. Abbott. Illuminating Engineering Society, New York, N. Y., Sept. 1953.

5. Peaking Ballast for Wide-Range Fluorescent Dimming. Arthur Bastian. *Electrical Manufacturing*, New York, N. Y., Aug. 1955.

A Brushless D-C Generator for Aircraft Use

J. T. DUANE
ASSOCIATE MEMBER AIEE

THE COMMUTATOR has long been a distinctive and sometimes troublesome feature of d-c generators. It functions as a mechanical rectifier converting the alternating armature voltages to direct voltage for the generator load. It is a well-known fact that the rectifying action of the commutator can be performed by static rectifiers if it is economically feasible to do so. In other words, a d-c generator can be replaced by an a-c machine with its output completely converted to direct current by static rectifiers. This has been done infrequently in the past.[1,2] Economics and application requirements have seldom justified the added complexity and increased cost. However, in modern aircraft applications a different situation exists. The increasingly severe environmental requirements placed on generating equipment have made it more and more difficult to obtain the necessary reliability and service life from commutator and brush systems. Higher altitudes have necessitated continual improvements in the ability of brush commutator systems to work under low-humidity, high-temperature conditions. In addition, demands for lighter weight generators have resulted in electrical designs which make satisfactory commutation even more difficult to obtain. The logical result of these trends is the elimination of mechanical commutators in certain air-borne d-c generating systems. This paper reports the development of such a system for a modern, high-performance intercepter aircraft.

The generating system described is rated at 100 amperes, 30 volts d-c. The generator requires no brushes and, in addition, the system eliminates the need for a carbon-pile type of regulator. Since the machine is designed for a high-performance aircraft, it is required to function in an exceedingly high-temperature environment. Generator cooling is accomplished by the circulation of oil rather than the use of blast air.

While the system described here is quite unusual in respect to most d-c systems

Paper 58-916, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing May 2, 1958.

J. T. Duane is with the General Electric Company, Erie, Pa.

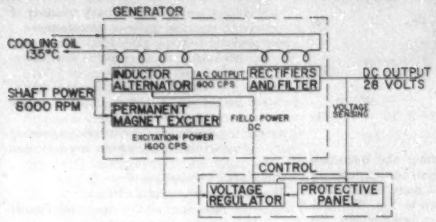


Fig. 1. Block diagram of brushless d-c generating system showing major components

Fig. 2. Completed generator and control panel

used in the past, it is quite typical of present developments in the aircraft generator field in at least two respects. First, generator and control components are required to operate under exceedingly severe environmental conditions. In addition to high ambient air and cooling oil temperatures, components are subjected to severe vibration and high accelerations. The second feature of the system development which seems typical concerns the fact that it is tailored to a specific aircraft application. It has been designed and built as one element of a complete weapons systems, not as a standard model available for a variety of applications. The design is matched to the space limitations of the installation and to the electrical and mechanical requirements of the aircraft.

This paper proposes to provide a brief description of the complete d-c generating system along with more detailed descriptions of various components. It relates some of the problems encountered during the development, and the solutions which were selected for these problems. Major emphasis is placed upon the discussion of the generator itself since many of the important features of the voltage regulator have been described in another paper.[3]

Generating System

The combination of components used to obtain brushless d-c generation is shown in the block diagram of Fig. 1. The elements shown are combined into two separate packages, as pictured in Fig. 2. The exciter, the main output generator, and the rectifiers are included in the generator itself with provisions for cooling by oil circulation. All voltage regulation and protective functions are incorporated in a separate control panel which requires no forced cooling.

Examination of Fig. 1 shows that an unusual approach has been adopted for the supply of excitation power to the main generator. Most d-c machines are self-excited by feeding back the d-c output through a variable resistance to the field winding. In most aircraft systems this variable resistance is a carbon-pile device which is sensitive to vibration and to the accelerations imposed by maneuvering and shock loads. Under some conditions this sensitivity can lead to instability and system malfunctions, a tendency to voltage drift with temperature variations or during warm-up is another undesirable characteristic of the carbon-pile regulator. For this development, an alternative approach was used. Excitation power is supplied through a magnetic-amplifier type of regulator, by a 1,600-cps (cycles per second) permanent-magnet generator (PMG) mounted on the main generator shaft. This scheme offers four important advantages:

1. The availability of an independent excitation source eliminates the need for the acceleration-sensitive carbon-pile regulator. A magnetic-amplifier regulator can be utilized to convert the 1,600-cps output to direct current for excitation purposes, and to provide control power at the same time.

2. The high-frequency PMG in combination with the magnetic-amplifier regulator permits an exceptionally fast system response as compared to more typical d-c systems.

3. The availability of a separate excitation source independent of generator line voltage makes it possible to supply high short-circuit currents without battery assistance. In other words, the system will not collapse the way a typical d-c system will under a short circuit. This improves the ability to clear system faults and can be an important factor in many aircraft applications.

4. Permanent-magnet excitation eliminates the need for the field flashing circuits often provided to insure generator build-up.

The main output generator is a double-ended, homopolar inductor alternator. The 12-pole construction provides an 800-cps output frequency at the rated speed of 8,000 rpm. The inductor design eliminates not only brushes and commutator, but also all rotating windings. This results in a rugged, highly reliable construction, and minimizes maintenance problems. The inductor approach has an added advantage in that it requires very low levels of excitation power, thus simplifying the excitation and control system.

The 800-cps output of the inductor alternator is converted to direct current by means of a 3-phase double-way rectifying system. The silicon rectifiers used are normally rated at 50 amperes, but they require a slight derating because of the high temperature of the cooling oil used. Filtering of the output to reduce audio-frequency ripple due to commutation of the rectifiers is accomplished with tantalum capacitors. The use of high-capacity tantalum units permits a much lighter weight system than would an inductive filter of comparable effectiveness.

The oil circulation system for the application supplies oil separately to the generator rotor and stator, thus eliminating the need for any high-pressure rotating oil seals. Oil enters the stator through the face of the mounting flange and is circulated first to the rectifiers, and then returned by a path which takes it around the outboard bearing and then through tubing on a series of passes around the stator frame to provide cooling there. From this point it is returned to the constant-speed drive gearbox. It is then supplied directly to a tube in the center of the rotor shaft which carries it down the length of the shaft to cool the inner race of the outboard bearing. In returning along the inside of the outer shaft, it provides the small amount of cooling necessary to remove stray pole face losses from the rotor structure. A small amount of oil is bled from the shaft just before it leaves the

machine to provide lubrication for the drive-end bearing.

In operation, the entire generator is pressurized to 7 to 9 pounds per square inch gage with purge air to prevent the flow of oil (particularly oil mist) from the constant speed drive back into the generator. To minimize the leakage of this purge air from the generator cavity to the drive, a clearance-type labyrinth seal is provided to limit air leakage in this area to 0.5 cubic foot per minute with an 8-psi pressure differential.

The components included in the control panel of Fig. 2 were developed and built by the General Electric Specialty Control Department in Waynesboro, Va. The voltage regulator is essentially the same as one developed for an a-c generator and previously described by Sechrist and Carlson.[3] The a-c regulator was redesigned for this application to use a Zener diode type of reference and a transistor preamplifier. This provides a transistorized and static voltage regulator which is not subject to the acceleration and vibration sensitivity problems sometimes associated with d-c regulators. Circuitry in the regulator compares generator output voltage with the Zener diode reference, and the resulting error signal is amplified in the transistor circuit. Transistor output establishes the current in the magnetic-amplifier control winding, which in turn controls the amount of PMG power fed through to the main generator field winding so as to maintain generator voltage constant.

Inductor Alternator

While guide specifications required a brushless generator, this in itself was not enough to specify the final generator configuration. There are many different generator types which might have been applied. One obvious approach would have been to use a permanent-magnet type of machine with voltage control by means of saturating windings on the stator. The major drawback to this type of generator was the fact that the excitation system serves to depress generator output voltage. In the event of any failure in the regulating system, the generator excitation would automatically go to ceiling, resulting in system overvoltages which could not be controlled except by a main line contactor. (No line contactor need be used with the present system.) In addition, this generator would deliver extremely large amounts of power into any internal faults with no means available for controlling or removing excitation to prevent the danger of fire. An asynchronous induction generator might also have been used, but this would have required complex voltage control circuitry. Another obvious possibility is the use of a conventional salient pole generator with a separate exciter and rotating rectifiers on the shaft to eliminate the need for brushes. This approach would have resulted in added complexity and a significant decrease in reliability because of the addition of two sets of rotating windings and an extra set of rectifiers in the rotor, where effective, reliable oil cooling is difficult to achieve. After reviewing the various possibilities, it was concluded that the best choice for the main element in a system of this type was an inductor alternator. The inductor alternator approach permitted the elimination of all rotating windings and rectifiers, and resulted in a more reliable and dependable system. Since no significant heat sources were to be located in the rotor, cooling problems in this area would be materially reduced, again increasing reliability. The inductor machine typically requires relatively low levels of excitation power. This again was a point in its favor since it permitted a low-weight exciter-regulator combination. There were, however, two different types of inductor alternators which might have been used. The heteropolar inductor machine uses multiple d-c field poles created by field windings lying in essentially the same axially-directed slots as the a-c output winding. The other approach was the homopolar type of machine where one coaxial field coil provides all of the excitation with no sacrifice of slot space in the stator. The homopolar approach proved the more effective since it permitted a lower weight generator for a given output. In addition, the homopolar approach made it simpler to provide a complete balance between the various phases of the a-c output. Phase balance can be achieved in the heteropolar device, but in some cases this requires a certain amount of sacrifice in weight.

The basic principle of the inductor alternator in all of its varied configurations is well described in the literature.[4–6] In spite of this, it should be helpful to describe briefly the way in which the double-ended, homopolar version used in this system functions. This can be done conveniently with reference to the conventional, salient-pole synchronous generator pictured in Fig. 3(A). The alternatively north and south magnetic poles on this rotor are induced by d-c field windings placed around each pole piece. Since these windings encircle each pole, they must rotate with the rotor, and therefore field current from a stationary source must be supplied through some sort of sliding contacts or brushes to the rotor. This conventional generator can be converted into a homopolar inductor machine by the changes pictured in Fig. 3(B). The rotor body and the stator are lengthened, but the pole pieces themselves are unchanged. The north poles are all placed at one end and the south poles at the other, but their angular relationship is unchanged. The modified rotor would still work with the original field coils, but this is unnecessary. All of the magnetic flux which passes from the south poles to the north poles internal to the rotor must go through the center portion of the shaft. A single, coaxial field coil around the shaft at this point will link all of the flux and supply all of the necessary excitation. This coil does not have to

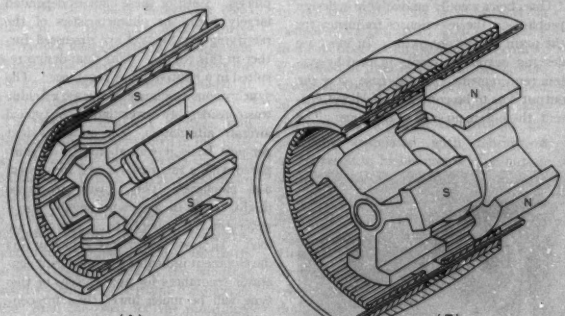


Fig. 3. Comparison of salient pole and homopolar inductor alternators

A—Conventional salient pole design B—Homopolar inductor machine

rotate and it can therefore be fixed in the stator, eliminating the need for brushes or other sliding contacts.

It is evident from Fig. 3(B) that the flux must travel axially both in the center of the rotor and outside the stator core. To accommodate this flux, solid steel paths are provided at both points. The coaxial coil is required to supply only as many ampere-turns as originally supplied by two of the separate coils of Fig. 3(A). If the generator has a large number of poles this can result in a significant reduction in excitation power. The sketches of Fig. 3 might suggest that the inductor machine is more than twice as large as the conventional machine. This is not the case. The inductor is usually heavier, but the typical increase is only about 60%. For some applications the lower excitation power requirements and elimination of current collection provisions will actually result in a reduction in system weight or volume.

The actual rotor developed for this application is shown in Fig. 4, along with the stator. The coaxial field coil can be seen inside the stator. The stagger in the rotor poles is provided to minimize slot ripple effects which would otherwise be reflected as ripple in the d-c output. The rotor teeth at the two ends are offset from each other by one-half pole pitch to obtain the 180-degree phase reversal necessary in the homopolar type of machine, since flux crosses the gap from rotor to stator at one end and from stator to rotor at the other. As may be seen from Fig. 4 one end of the rotor must be arranged so that the punchings may be removed for assembly in the stator. Since the rotor punchings are keyed on the shaft, this requires only that a single nut be removed and the punching stack slipped off.

One choice usually unavailable in design problems is the selection of frequency for the main output generator. In most a-c designs, output frequency is fixed by system requirements. In this case, since the output was to be converted to direct current, there were no such restrictions. The designer was free to select frequency (number of poles) in a manner which would provide the lowest system weight. In making the choice a number of factors were considered. For fixed output, generator weight will vary significantly as the number of poles or frequency is changed. In addition, since end-turn extension will vary greatly as pole span changes, the selection of the number of poles has a marked effect on the over-all density of the main generator. This can be an important factor in instances where available space is limited. Another factor not directly connected with the generator weight is the effect of frequency on the rectifying and filtering system. The rectifiers themselves see no frequency limitations within the range of feasible frequencies. However, filtering of the output ripple to provide high-quality d-c output does depend greatly on frequency. Requirements for filtering favor the highest possible generator frequency. The ripple frequency for the rectifying system used here will be six times the a-c output frequency of the alternator. The limitations imposed by the variation in generator weight-per-unit output as a function of the number of poles are not too severe in this particular case. Inductor alternators seem to have a fairly broad minimum weight region when weight per unit output is plotted against the number of poles. Consideration of all these factors led to the design of the inductor for this application as a 12-pole unit providing 800 cps at its 8,000-rpm operating speed.

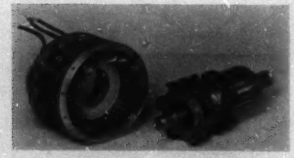


Fig. 4. Disassembled rotor and stator showing inductor construction with coaxial field coil visible in stator

Design of the generator to obtain the proper reactances for optimum system weight and performance presented other interesting problems. One factor not usually available to the designer as a variable was the choice of the number of phases. Both of these choices depended largely upon the characteristics of the rectifying system and are discussed further in this respect. The final design resulted in a relatively low reactance. The synchronous reactance of the a-c machine was considerably lower than most typical aircraft alternators. On an equivalent per-unit basis it was about 0.5 as compared with a range of 1.2 to 2 for typical a-c machines. This lower reactance might result in a weight penalty for a generator operating into a straight a-c load. However, experience gained in this development indicates that optimum generator reactances for applications of this type will be much lower than for conventional systems.

The electrical characteristics of the complete generator (with rectification) are not significantly different from a conventional d-c machine. Load saturation curves at several different line currents are shown in Fig. 5. One peculiar characteristic of inductor machines is that, as field current is increased, the slope of the saturation curve passes through zero and becomes negative.[6] In other words, terminal voltage starts to decrease with increasing field current above a certain point. This will occur in the subject generator, but the critical field current point is well above the ceiling power capabilities of the excitation system.

Self-excited d-c generators will build up to rated voltage if the resistance in the field circuit is held below a critical value determined by the connected load and the shape of the saturation curve. There was some question as to whether an a-c machine operating with silicon rectifiers would build up in a comparable manner if it were self-excited instead of being supplied from a separate exciter. To satisfy curiosity on this point, this inductor machine was tested in a self-excited connection with an aircraft-type carbon-pile regulator. The regulator conformed to military specifications for aircraft systems.[7] This combination built up to rated voltage with no difficulty under full load at minimum rated speed.

Rectification

Development of the rectifying system for this application presented several interesting problems. The major limitation was in the variety of available silicon rectifiers capable of operating at the required temperatures. Work in semi-

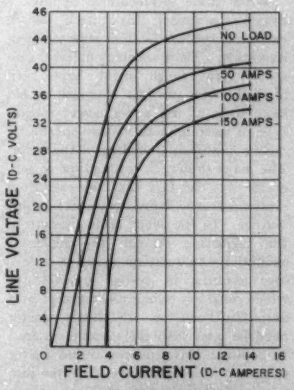


Fig. 5. Saturation characteristics of rectified output inductor machine

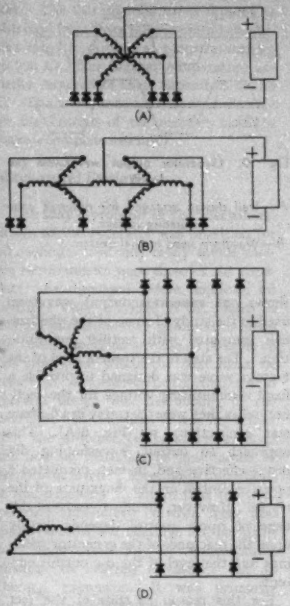


Fig. 6. Possible rectifier connections

A—Six-phase star
B—Double Y with interphase
C—Five-phase full-wave bridge
D—Three-phase double way

conductors has been progressing rapidly in recent years, but at the time that the final configuration had to be fixed, the number of available rectifiers was quite limited. An additional limitation was imposed by the fact that the cooling medium was at times only 50 C (degrees centigrade) lower in temperature than the maximum junction temperature at which most of the rectifiers could be operated. There are many rectifier combinations which might have been used in conjunction with various alternator-winding arrangements. Some of these are illustrated in Fig. 6.

The 6-phase star configuration shown in Fig. 6(A) was one arrangement which would obtain maximum current output from available rectifiers. This was an important factor since, in a 30-volt system, rectifier voltage limitations are not serious. The fact that all six of the rectifiers conduct in parallel permits the output current limit to be almost, but not quite, six times the average current rating of any one of the individual cells. This system did have two drawbacks. The most serious of these was the fact that it requires an appreciably larger generator. This was partly due to the narrow current pulse associated with the 60-electrical-degree conduction periods of the rectifiers. The resulting current waveform led to increased heating and demagnetizing effects in the generator. The stator connection also led to relatively high winding reactances which in turn introduce two additional problems:

1. High reactances introduce regulation effects in the rectifying circuit which increase generator size.

2. High reactance makes for much poorer commutation in the rectifiers and consequently higher levels of ripple voltage in the d-c output. This higher ripple would require a larger, more expensive filtering system, and it proved to be a major factor in ruling out the 6-phase star connection.

It seems appropriate to consider at this point the commutation effects in rectifier circuits. In the usual d-c machine, current in an individual coil is commutated, or undergoes a reversal as its commutator segments pass under the brushes. The relative motion of brush and segment force current reversal in a fixed interval (with fixed speed) independent of current magnitude or winding reactance. An analogous but not identical phenomenon occurs when static rectifiers are used to convert polyphase alternating current to direct current. Since current flow in one phase represents energy stored in a magnetic field, that current cannot switch instantaneously to the next phase. A finite time is required for the current transfer to take place. This results in some overlap in the conduction periods of adjacent phases, and a drop in terminal voltage corresponding to the average voltage of the two phases undergoing commutation. Commutation effects, therefore, reduce average output voltage and increase ripple voltage. These effects are well known and they are discussed at length in the technical literature.[8,9]

The 3-phase double-Y system with interphase transformer shown in Fig. 6(B) was a possibility which would minimize commutation effects. Again, maximum rectifier current capacity is obtained, but an interphase transformer is required. Generator weight must be somewhat higher than minimum because of the increase in copper losses in the windings caused by half-wave current conduction. Other than the need for the interphase transformer, this system is markedly superior to the 6-phase star. Generator output power requirements are not as high, and with a properly designed transformer the ripple in the d-c output is approximately equal to that obtained with a 3-phase double-way connection.

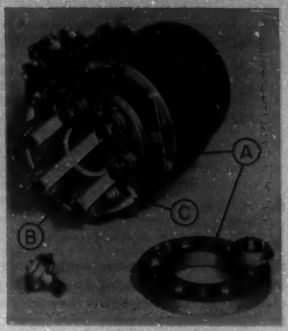


Fig. 7. Rectifier end of generator with cover removed

A—Negative cooling plate, grounded
B—Positive cooling plate
C—Teflon gasket

Fig. 6(C) shows another configuration which is typical of some of the many possibilities which were available. The system shown uses five phases with full-wave rectification, but almost any number of phases might have been chosen. The major problem in selecting any of these other possibilities was that of properly matching the rectifier capabilities to system requirements. If, for example, the available rectifiers were capable of performing properly in the 3-phase double-way system of Fig. 6(D), the same rectifiers applied in the 5-phase system of Fig. 6(C) would have been misapplied since they would be larger than necessary to do the job. This would result in excessive weight, cost, and complexity in the rectifier-cooling system. Of course, if rectifiers of three fifths of the rating were available, they might properly be applied in the 5-phase system. Although material cost might still be comparable, the problems associated with fabrication would result in higher cost and perhaps a slightly lower reliability. This 5-phase system would result in a higher ripple frequency and a lower magnitude of ripple, thus simplifying filtering problems. However, the major limitation proved to be the fact that rectifiers were available which were suitable for the 3-phase double-way system and there were no lower rated rectifiers which could be used with comparable effectiveness in a 5-phase system.

The rectifier connection finally chosen

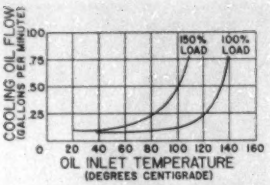


Fig. 8. Generator oil cooling requirements

was the 3-phase, double-way arrangement of Fig. 6(D). This circuit does not require any more rectifiers than the others, but the individual cells must have a higher current rating. It minimizes generator size since it causes the lowest winding losses and demagnetizing effects of any of the other three (or six) phase schemes. The double-way connection provides minimum peak inverse voltage on the rectifiers, but it doubles the forward voltage drop which the generator sees in the rectifier.

This discussion of rectifier circuits has intentionally avoided the quantitative figures of merit commonly associated with transformer rectifier systems used in industrial applications. No attempt has been made to apply such terms as power factor, secondary volt-amperes, per-cent regulation, or efficiency to the rectifier circuits considered. This has been done on the basis of the author's conviction that it would be misleading to try to apply conventional criteria without any modification to the problem of designing a co-ordinated generator-rectifier system. There surely is much that can be very useful in available rectifier theory. However, the design problem considered here is sufficiently unusual to require a careful evaluation on its own merits before existing theories are applied. There is much work yet to be done in this area and it should be a fertile field for future investigation.

The completed rectifier assembly is shown in Fig. 7 with the generator cover removed. Cooling is accomplished by removing heat from the large stud at the bottom of the individual rectifier cells. The rectifier junction is limited to operating temperatures of approximately 200 C and the thermal configuration of the stud and junction is such that a temperature rise of approximately 0.8 C for every watt dissipated is necessary between the stud and the junction. Since these rectifiers have losses of approximately 35 watts each at full load, the generator must be installed and cooled in such a way that the stud temperature is maintained at a maximum of 172 C. Actual operating stud temperatures are somewhat lower to allow for transient overcurrent conditions. A further complication is introduced because requirements for integrity of the oil system prevent direct contact between cooling oil and rectifier stud. This cooling problem was solved by the use of the rectifier cooling plates visible in Fig. 7. Two plates provide paths for both current and heat flow from the rectifiers. Oil is circulated within the plates around the base of each of the individual rectifiers to provide maximum cooling. This configuration proved extremely effective, with a maximum temperature differential between inlet oil and the top of the rectifier stud of approximately 15 C under rated cooling and load conditions. Oil flow between the rectifier plates and the generator end shield is accomplished by means of holes in the face of the plates encircled by O-rings providing the necessary hydraulic seal. Thus electrical isolation of the plates is possible. In actual practice only one of the rectifier plates is isolated from its mounting surface. The negative side of the generator output is directly grounded internal to the generator. The second plate, corresponding to the positive terminal of the generator, is of course isolated. It is the outer half-plate of the two shown in Fig. 7. Electrical insulation between the two plates is provided by a sheet of Teflon, which establishes the necessary barrier to current flow but does not in itself perform any sealing function. The Teflon does provide an interface against which O-rings in each of the two adjacent plates are pressed to provide the necessary oil seals. The Teflon gasket serves a second function as a supporting structure for leads associated with the filter assembly.

The effectiveness of the rectifier-cooling provisions are illustrated by the rating curves of Fig. 8. Minimum allowable cooling oil flow through the generator is shown as a function of oil inlet temperature. Rectifier temperatures place the major restriction on rating since specification restrictions on efficiency precluded a generator design with losses high enough to make the windings thermally limiting. It is evident from Fig. 8 that the oil flow versus inlet temperature curves are asymptotic to a vertical lines, indicating maximum allowable rectifier plate temperatures.

In the development of this generator, many different approaches were considered and evaluated for the reduction of the audiofrequency ripple in the d-c output voltage. Considerable work was done on the reduction of ripple by shaping of the generator output waveform. Fig. 9(A)

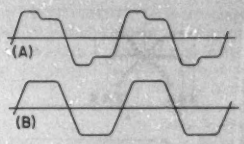


Fig. 9. Generator output waveforms (no load)

A—Trail design arranged for optimum commutating voltage

B—Waveform used in final design

shows one generator output waveform used in the study of some of the phenomena associated with rectifier commutation. The step in the trailing edge of the voltage wave was designed to provide a fixed commutating voltage for the rectifiers when they were operating in a 6-phase star connection; see Fig. 6(A). This approach to output waveshaping did prove effective and, in fact, permitted a positive control on the magnitude of the ripple. However, the effectiveness of this type of ripple control depends largely upon the reactance of the generator windings and the level of the d-c output current.

For this reason, in spite of the fact that waveshaping did prove effective, it was not used because it would have required an excessively low reactance (a very large generator) to make it effective at the current levels required for this application. The waveform of Fig. 9(B) was selected because it permitted maximum generator output. Another approach to the elimination of ripple is to provide brute-force filtering. Most rectifier applications use large inductors or chokes in the d-c line to provide a high reactance or impedance to ripple currents so that the ripple voltage appears entirely across the inductor. While several different inductor configurations were considered and evaluated for this application, they did not prove suitable because of the weight required. The problem was further complicated by the fact that since the inductor had to carry full d-c load current, it was difficult to cool properly. A much more effective and economical solution from a weight standpoint was the inclusion of capacitors to provide a low impendance shunt path for the high frequency (4,800 cps) ripple so that only a small portion of it appeared across the load. The final configuration uses 240 microfarads of capacitance directly across the d-c output. This type of filter was made possible by recent developments in small, high-temperature tantalum capaci-

tors. The 240 microfarads of filtering is obtained from three 80-microfarad capacitors rated to operate at temperatures up to 200 C. The highest temperature actually required of these capacitors is 155 C. The capacitors are mounted directly on the bottom of the rectifier plate to make a complete assembly.

Mechanical Design

The major problem in the mechanical design of the generator was provision for oil cooling. The only really severe cooling requirements were those of the rectifiers. As discussed before, intricate oil passages were provided in the plates in which the rectifiers are mounted to bring the cooling medium as close as possible to the stud from which the heat must be removed. Another area requiring careful thermal design was the cooling of the bearings. In this case, it was important that the absolute temperature levels of the bearing materials be maintained at acceptable levels. Even more important, it was necessary to insure that the temperature differential between inner and outer races of the bearings did not grow too large, resulting in unacceptable changes in bearing clearances due to thermal expansion. It was primarily for this reason that oil was provided in the generator rotor so that the inner races could be maintained at approximately the same temperature as the outer races.

Cooling of the main generator windings is accomplished through the oil tubes on the outside of the stator frame shown in Fig. 2. Heat from the windings themselves is carried through the slot insulation and the stator iron out to the tubes. In some oil-cooled generators, it is necessary to bring the cooling medium much closer to the current-carrying conductors themselves; however, this was not necessary. The major reason that cooling was required only on the outside of the stator frame was the fact that this generator design is to a certain extent efficiency-limited. Specifications required a system efficiency of 75%. In a 3-kw generator this means that total losses under full-load conditions cannot exceed 1,000 watts. Holding losses to this level is complicated by the fact that essentially fixed losses, such as windage friction and excitation, account for a significant percentage of the total allowable loss. The most difficult of the windings to cool effectively was the coaxial field coil. Heat from this coil is transmitted through the coil bundle to its outer surface. From this point on, it was possible to take special precautions to provide a thermally efficient conduction path to transmit the heat from the outer surface of the coil to the stator frame where it can be removed by the cooling oil tubes.

The relatively high speeds and very high temperatures required for the bearings in this unit again required care in mechanical design. To eliminate any rotating seals, the outboard bearing is grease-lubricated. This bearing, therefore, required special heat-stabilized bearing steels with high-temperature grease to provide the necessary service life. This, in conjunction with careful provisions for cooling of both inner and outer race have provided satisfactory bearing performance. The drive-end bearing presented a simpler problem. It was entirely compatible with the application to provide direct oil lubrication of the bearing. This proved very fortunate since this bearing is required to carry all thrust load on the shaft and the hydraulic configuration of the drive and cooling system requires that the bearing carry approximately 200 pounds of axial thrust load. To accommodate this loading, special bearing clearances were provided to maximize thrust load capacity.

Conclusions

The analytical studies, design work, and development effort reported here have led to several conclusions:

1. The design of these systems requires careful matching of possible generator and rectifier configurations. While some headway has been made in formulating standards and criteria applicable to the problem, there is much work to be done before understanding of these systems is brought to the point already reached in other machinery design areas.

2. Capacitor filters offer significant advantage in both weight and simplicity for the elimination of ripple in rectified output a-c machines.

3. The use of permanent magnet generator excitation leads to improved system performance and helps to eliminate undesirable system components.

4. The inductor alternator offers major advantages in simplicity of design, ruggedness of construction, and ease of maintenance. Its adaptability to severe environmental conditions also helps to make it particulary suitable for aircraft-type applications.

5. Brushless d-c generators are not only possible, but practical, for some aircraft applications. Systems of the type described here should see continued use in the future as environmental requirements become increasingly severe.

References

1. Selenium Rectifiers in Motor Vehicle Power Systems, Glen Ramsey. *AIEE Transactions*, vol. 68, 1949, pp. 149–53.

2. An A-C Vehicle Generating System, A. D. Gilchrist. *Electrical Engineering*, vol. 66, Aug. 1947, pp. 779–80.

3. Regulator-Exciter Systems for Small Missile and Aircraft Power Supplies, H. O. Carlson, H. S. Sechrist. *AIEE DP57–525* (available on request).

4. The Theory of the Inductor Alternator, J. H. Walker. *Journal*, Institution of Electrical Engineers, London, England, vol. 89, pt. 2, 1942, pp. 227–41.

5. Short Circuit Characteristics and Load Performance of Inductor Alternators, A. Mandl. *Ibid.*, vol. 94, 1947, pp. 102–17.

6. High-Frequency Alternators, J. H. Walker. *Ibid.*, vol. 93, pt. 2, 1947, pp. 67–80.

7. General Specification for Aircraft Engine Driven 30-Volt Direct Current Generators. *Military Specification MIL-G-6162*, Sept. 1947.

8. Analysis of Rectifier Circuits, E. F. Christensen, C. H. Willis, C. C. Herskind. *AIEE Transactions*, vol. 63, 1944, pp. 1048–58.

9. Principles of Mercury Arc-Rectifiers and Their Circuits (book), D. C. Prince, F. B. Vodges. McGraw-Hill Book Company, Inc., New York, N.Y., 1927.

Discussion

R. K. Walter (Convair, San Diego, Calif.): The air frame manufacturer's specification for this system required a brushless d-c generator with static regulation to meet the extreme-altitude, speed, and temperature requirements of the aircraft This has provided for the aircraft electric system designer a solution to the many disadvantages associated with the conventional mechanical rectifier. Some of these disadvantages are that it is a source of heat; an explosion hazard due to sparking; difficult and expensive to manufacture; a variable source of radio noise; there is limitation in its speed of rotation; and there is the ever-present question of brush life under all operating conditions. The static rectifier approach, with inherent long life, has left primarily the bearings as a maintenance and overhaul item. This machine is one of four generators comprising the electric system of the *F-106* interceptor. Two generators supply six special voltages and frequencies for a fire control system, while the remaining two generators provide for general-purpose aircraft power.

The inductor generator lends itself to high-speed high-frequency application; thereby providing promise for future generating systems where solid rotor design is necessary because of temperature limitations.

During the development of the generator, the following problem areas were of greatest concern to assure compatibility between the generator and drive:

1. The nonstandard generator mounting pad required close co-ordination effort between the vendors concerned and the air frame manufacturer.

2. The unconventional method of driving

the generator using a gear, which at the same time provides a sealing surface to capture return oil from the drive, along with the inlet rotor oil tube seal, required extremely close tolerances to be held by each manufacturer. The gear drive permits a weight reduction of two bearings since the generator bearings serve to support the rotor and the drive gear.

3. Maintaining purge air pressure integrity of the generator housing with relatively small quantities of air available required the development of special sealing compounds to achieve airtight seals. A generator design change is now pending, incorporating the ability of the generator to operate without purge air pressure by using a face-type carbon seal and an internal oil drain. In addition, by lubricating the generator drive-end bearing with drive splash oil, the bearing lube hole has been eliminated. These mechanical design changes should result in greater ease of manufacture, better quality depot repairability, and longer in-service life.

Supplementing Mr. Duane's comment regarding the fact that this system functions without the use of a line contactor, it is possible to operate in this manner since all switching functions, (i.e., over- and under-voltage and underfrequency, pilot's switches and external power interlock), perform the operation of either turning the system on or off. With the system turned off, the very small residual generator voltage permits personnel to perform maintenance operations without hazard.

In addition, operating without a line contactor allows the location of the voltage-sensing lead to be at the load center of the aircraft without the use of a voltage-sensing transfer relay. This relatively long wire length uses no fuses; therefore mechanical protection has been afforded by using a shielded wire with an extruded Teflon jacket in addition to careful wire routing.

Experience has shown that the PMG working into transformer rectifier loads in the control panel produces radiated noise in excess of the limits prescribed in Military Specification *MIL-I-6181B*. The solution to this problem was achieved by the use of enclosing case construction for the control panel (in lieu of a potted block exterior) and the use of shielded wires for the PMG output and generator field leads. The advantage gained is added physical protection for these critical circuits.

In summary, experience has indicated that the brushless d-c generator and static control panel system are well suited to high-performance aircraft application.

J. T. Duane: Mr. Walter's comments are very much appreciated. His remarks concerning features of the aircraft installation should help to clarify the manner in which the generator has been applied.

External Heat Transfer from Cylindrical Self-Cooled Electric Equipment

DANIEL FRIEDMAN
ASSOCIATE MEMBER AIEE

IN ADDITION to the usual electromagnetic and electrical limitations, the performance of electric equipment is dependent on its ability to dissipate its losses and still maintain safe component temperatures. For forced-cooled equipment, such as blast-cooled generators, a continuous flow of coolant is provided to carry away practically all of the heat. Since only one method of heat transfer need be considered, techniques are available which can be used to predict the cooling performance for any set of cooling conditions encountered in aircraft.[1] For self-cooled equipment, however, the heat transfer picture is much more complicated. Heat may be dissipated by natural or forced convection to the surrounding air, by radiation to surrounding objects, and by direct conduction to objects with which it comes in contact.

The importance of the external air velocity and the radiation environment has not always been fully appreciated. As a result, the tacit assumption is frequently made that the specification of air temperature and pressure is sufficient to define the cooling environment. It was found, however, that supposedly identical tests run in a number of environmental facilities were surprisingly different. This discrepancy is the result of the fact that different types of environmental facilities provide different air velocities and radiation environments for the same air temperature and pressure. For example, an old chamber with a high air leak rate will usually have a higher air velocity than a tight one.

In order to get reliable and reproducible data, two approaches have been developed. The Naval Air Test Center (NATC) uses a perforated chamber mounted in the center of the main altitude chamber. Fig. 1 shows a typical installation. This enclosure provides a standardized set of environmental conditions which are effectively independent of the characteristics of the particular chamber. Some of the results of the

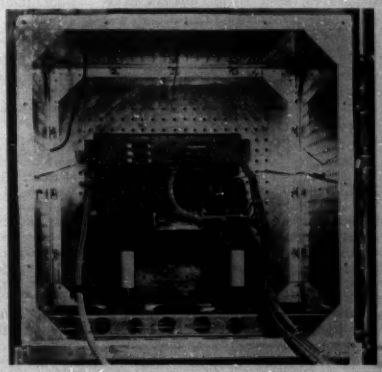


Fig. 1. Perforated enclosure for use in the thermal evaluation of self-cooled equipment

Paper 58-857, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 13, 1958; made available for printing April 11, 1958.

DANIEL FRIEDMAN is with the U. S. Naval Research Laboratory, Washington, D. C.

The author wishes to acknowledge the contributions of members of the Self-Cooled Group, Rating of Airborne Electrical Apparatus Subcommittee, AIEE Air Transportation Committee. The design of the drum installation was based on guidance obtained at meetings of the Self-Cooled Group. In addition, some ideas presented in this paper resulted from the discussion held at these meetings.

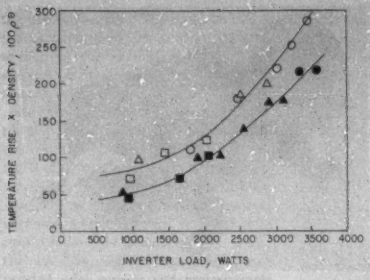
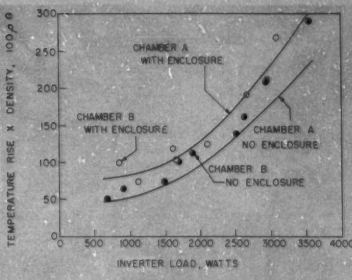


Fig. 2. Correlation of inverter frame temperature measurements, chamber A

○ 20,000 ft, with enclosure
● 20,000 ft, no enclosure
△ 35,000 ft, with enclosure
▲ 35,000 ft, no enclosure
□ 50,000 ft, with enclosure
■ 50,000 ft, no enclosure

Fig. 3 (right). Comparison of inverter frame temperature measurements


TEMPERATURE RISE X DENSITY, 100ρθ
CHAMBER A WITH ENCLOSURE
CHAMBER B WITH ENCLOSURE
CHAMBER A NO ENCLOSURE
CHAMBER B NO ENCLOSURE
INVERTER LOAD, WATTS

NATC study will be discussed in the next section of this paper. The other approach is to construct a test facility in which all the important factors can be controlled. The design and use of such a facility is discussed in a later section of this paper.

In order to get a better understanding of the phenomena involved, a study was made of the basic modes of heat transfer using classical theory and an experimental evaluation made to determine how well the theoretical results can be applied to predict the characteristics of a typical piece of equipment. This paper contains a discussion of the results of this program.

Perforated Enclosure

In order to evaluate the effectiveness of the perforated enclosure, thermal tests were run at NATC in two different types of chambers. The enclosure used, see Fig. 1, was a 2 by 2 by 2-ft (foot) box with 3/8-in. (inch) holes on 2-in. centers. The frame temperature correlation shown in Figs. 2 and 3 is typical of the results obtained. In presenting the data, the first step was to determine a presentation method which would reduce or, if possible eliminate the effect of altitude. Utilizing a method originally proposed by Franklin,[3] correlation was tried using temperature rise times $(\text{density})^{1.0}$, and temperature rise times $(\text{density})^{0.9}$. The unity power value was used since it seemed to give the best results.

In both chambers, the use of the enclosure resulted in higher frame temperatures. This is to be expected since the enclosure reduced the air velocity at the surface of the unit. It should be noted, however, that the effect of the enclosure was larger in the chamber *A* tests. This can be explained by the fact that the construction of chamber *A* was such that for the same altitude and air temperature it would have a higher air velocity. Fig. 3 shows that the use of the enclosure will yield test results which are substantially independent of the chamber characteristics. The results of a complete thermal evaluation using the perforated enclosure are contained in reference 3.

Drum Configuration

The perforated enclosure is a useful device for qualification and comparative evaluation purposes. It cannot be used to study the basic processes involved in the external heat transfer from self-cooled equipment because:

1. The magnitude and direction of the external air velocity is unknown. Although devices are available which can be used to measure these variables, it is believed that the complexity of the air flow patterns within the enclosure would make analysis of the results extremely difficult.

2. The air velocity and radiation environment cannot be varied. This control is necessary if it is desired to separate the effects of the different environmental factors.

To provide the required environmental control, the configuration shown in Figs. 4 and 5 was constructed. The required air flow is supplied by a variable-speed supercharger. The total quantity of air flow is measured by an orifice installation designed according to American Society of Mechanical Engineers standards[4] and the average air velocity calculated using equation 15. Two inlet screens and the narrow angle diffuser are used to insure uniform air velocity in the drum, and pressure and air temperature are controlled by locating the drum and air flow equipment in an altitude chamber.

The electrically heated model shown in Figs. 6 to 9 is used to simulate a typical self-cooled electric device. This type of device has the following advantages:

1. The magnitude and distribution of the internal heat flow is fixed. One of the difficulties with the use of a rotating electric device for heat transfer studies is the fact that the loss characteristics are a function of altitude and show considerable variation with time. This variation in characteristics accounts, in part, for the scatter in data shown in Figs. 2 and 3.

2. Since there is no electric or mechanical output, the installation and instrumentation problems are reduced to a minimum. For example, the use of a motor would require a means of dissipating the output power as well as torque and speed measuring equipment. The heated model can, therefore, easily be used as a calibrated device to determine environmental conditions in an aircraft compartment under flight conditions. It can also be used to calibrate existing altitude chamber facilities.

3. It is a rugged device which can easily be duplicated in any machine shop. Detailed construction drawings can be obtained by writing to the author.

4. The thick brass walls and large internal

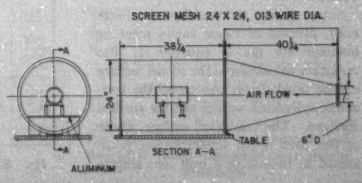


Fig. 4. Drum configuration and installation of test model

Interior and components painted with high-temperature standard black paint

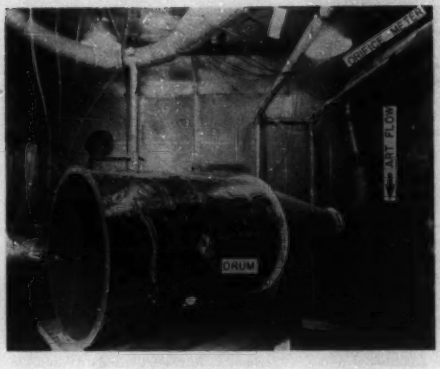


Fig. 5 (left). Drum installed in altitude chamber

Fig. 6 (above). Test model

heater reduce the model surface temperature gradients to a minimum. This keeps the problem of calculating the mean or average surface temperature as simple as possible.

Theoretical Analysis

The external heat transfer from self-cooled equipment involves three basic methods of heat transfer; namely, convection, radiation, and conduction. For these studies it is assumed that the direct conduction of heat is negligible. During the experimental phase of this study, the model was mounted on insulating rods in order to insure that the conductive heat transfer would be negligible.

The radiant heat transfer can be calculated using the Stefan-Boltzmann law, namely

$$q = \frac{3.69}{10^8}\epsilon_t\epsilon_m a\left[\left[\frac{T_s}{100}\right]^4 - \left[\frac{T_r}{100}\right]^4\right] \qquad (1)$$

Using the emissivity values for flat black lacquer contained in Table XIII of reference 5, it was assumed that $\epsilon_t\epsilon_m = 0.95$. The determination of the effective area for radiative heat transfer is complicated by the fact that the legs intercept some of the radiation from each other and from the main cylinder. To allow for this effect, it was assumed that the effective radiation area is equal to the area of the cyclinder plus 1/2 of the surface area of the legs.

Since surface temperature gradients exist, average or mean values must be obtained for use in equation 1. For radiation calculations, the model surface can be divided into two parts: the horizontal cylindrical surface and the vertical surfaces. The cylindrical surface, whose surface area is 204 square inches, is assumed to radiate to a wall temperature equal to the average temperature at the center of the drum. The vertical surface, whose effective surface area is 60 square inches, is assumed to radiate to a wall temperature equal to the average of the air temperature and the chamber temperature. The formulas used to calculate the average temperature are listed in the Appendix. It should be noted that the mean of the temperature to the fourth power should be used for radiation calculations. However, since the temperature gradients are small, the arithmetic mean should be adequate.

The calculation of the convective heat transfer is a little more complicated. For convection in still air, the analysis of the natural convection from infinite horizontal cylinders and spheres contained in Section 25-1 of reference 6 was used. Converted to the units used in this paper, the correlation contained in reference 6 can be represented by the equation,

$$h = \frac{1.05}{10^3}\frac{\sqrt{P}\,\beta^{1/4}}{D^{1/4}\,T_f^{0.37}} + \frac{5.67}{10^7}\frac{T_f^{0.743}}{D} \qquad (2)$$

A detailed derivation of equation 2 is contained in the Appendix.

For forced convection, the analysis of the heat transfer between spheres and air contained in Fig. 122 of reference 5 was used. Converted to the units used in this paper, the correlation contained in reference 5 can be representedby the equation,

$$h = \frac{1.02}{10^4}\frac{(PV)^{0.8}}{T_f^{0.293}\,D^{0.4}} \qquad (3)$$

A detailed derivation of equation 3 is contained in the Appendix.

For convective heat transfer, it is assumed that the entire surface area, 280 square inches, has the same effective heat

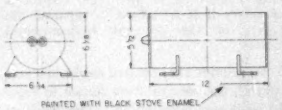


Fig. 7. Assembly drawing of test model

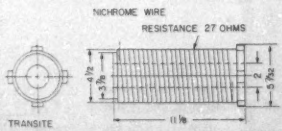


Fig. 8. Heater for test model

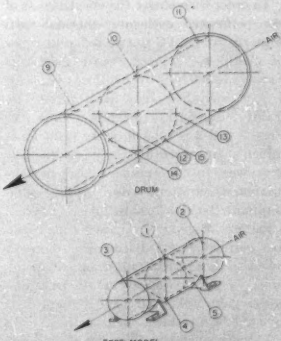


Fig. 9. Location of thermocouples

transfer coefficient. The mean temperature used for convection calculations is the average of the vertical and cylindrical surfaces weighted according to the relative areas involved (equation 29).

In studying forced convection heat transfer, there is an additional factor which should be considered. Reference 7 shows that the effective heat transfer from a cylinder, for a given air flow and pressure level, increases as the turbulence level increases. Although no measure was made of the level obtained in this study, the air flow configuration used incorporated inlet screens and narrow angle diffusers in order to keep the turbulence level below what might be expected in altitude chambers and aircraft compartments. For most applications, therefore, the results obtained in this study will be the conservative side.

In contrast to the round surface of the models used in the heat transfer studies correlated by equation 3, the model used in this study has sharp corners and flat surfaces. The test results for tubes with noncircular cross sections, summarized in Table 26-4 of reference 6, show that sharp corners and flat surfaces tend to increase the numberical value of the heat transfer coefficient and the exponent of the PV term. One objective of the experimental phase of this study is to determine the magnitude of this effect.

Experimental Results

Radiation

To check the validity of the assumptions used to calculate the radiative heat transfer, the test model was run in a large cylindrical tank, 50 inches in diameter and 7 feet long, at an ambient pressure of about 1.5 microns of mercury. For this configuration, see test no. 1 of Table I, the calculated radiation was 244 watts and the measured heat input was 249 watts. This is close enough agreement for the purposes of this study.

Natural Convection

Table I, test numbers 2 and 3, summarizes the results of measurements made under natural convection cooling conditions (no drum air flow). The convective heat transfer is assumed equal to the heat input to the heater minus the calculated radiative heat transfer. For the two tests listed, the average value of R is 1.26. The fact that R is greater than 1.00 may be the result of: 1. flow-disturbances resulting from the sharp corners and legs, and/or 2. air currents resulting from the realtively small diameter outer drum.

Multiplying equation 2 by 1.26 and using the physical constants for this test model, the equation for natural convection heat transfer becomes.

$$q = \frac{0.393\sqrt{P}^{0.55}}{T_f^{0.370}} + \frac{2.54}{10^4}\theta T_f^{0.743} \qquad (4)$$

Forced Convection

Table I also contains the results of tests made under forced convection cooling conditions. In Fig. 10, the measured heat transfer coefficients are plotted as a function of the theoretical values calculated using equation 3. As predicted, the results shown in Fig. 10 indicate that the actual heat transfer is greater than the theoretical value and that the exponent of the PV term is greater than 0.60. Incorporating the results of Fig. 10 and the physical constants for this test model, equation 3 becomes,

$$q = \frac{3.180(PV)^{0.65}}{100T_f^{0.393}} \qquad (5)$$

To determine the effect of orientation, test number 9 was run with the axis of the model at right angles to the axis of drum. In this case, the heat transfer coefficient was 10[1]/2% greater than the value which would have been obtained with coaxial orientation.

To check the effect of a change in drum temperature, test number 5 was run with the drum electrically heated. There was no significant change in the convective heat transfer coefficient from the results of an identical test (no. 4) run with an unheated drum.

Conduction

To obtain a rough measure of the effect of mounting conduction, test no. 10 was run with brass rods instead of plastic rods. In this case, the heat conduction through the legs was 26 watts or less than 6% of the total heat dissipation. This test would seem to support, for one type of installation at least, the assumption that the direct conduction of heat can be neglected.

Conclusions

1. For radiative heat transfer, the theoretical and measured heat transfer rates show very close agreement.

2. Under natural convection conditions, the heat transfer coefficients measured on the test model were greater than the theoretical values by about 26%.

3. For forced convection conditions, the heat transfer coefficients measured on

TABLE I RESULTS OF EXPERIMENTAL EVALUATION OF TEST MODEL

Test number			1	2	3	4	5	6	7	8	9^2	10^4
Absolute pressure, cm mercury			1.5×10^{-3}	76.1	3.78	76.02	76.23	8.64	8.73	8.69	76.24	76.33
Thermocouple reading, °F	model	1	170.7	164	181	135.5	140	122	198	189	129.5	132
	surface	2	151.7	158	163.5	102.5	106	85.5	167.5	157.7	100.5	99
		3	146.7	171	172	133	137	119.5	189.5	183	123.5	127
		4	163.4	172	177	129.5	134	119.5	193	185	122.0	123
		5	162.8	161	168.5	111.0	115	99.0	176.5	168	107.0	106
	chamber			35	28	34	34.5	-40.5	31	39	33	32.5
	drum	9		48	54	38	55.5	-32	45	53.5	37.5	27.0
		10		55	66	39	58	-30	53	59	39	39
		11		48	53	38	42	-35	40	51	37.5	37
		12		50	63	40	62	-26	56.5	61	39.5	40
		13		50	62.5	39.5	58	-29	54	59.5	38.5	39
		14		48	60.5	39.5	60	-29	54	59	38.5	39
		15		48	61.5	39.5	57	-30	53	59	38.5	39
Total heat loss, watts			249	450	300	450	450	450	450	450	450	450
Radiation, watts			244	262	256	148	139	186	313	282	134	136
Conduction, watts			0	0	0	0	0	0	0	0	0	26
Convection, watts			0	188	44	302	311	264	137	168	316	288[3]
Air velocity, ft/min			vacuum	natural		245	242	717	286	572	239	243
PV				convection		18500	18450	6180	2500	4970	18230	18550
Heat transfer coefficient		calc.		.00378	.0091	.0080	.00800	.00429	.00235	.00355	.0121[3]	.01213
		measured		.00498	.0110	.0123	.01205	.00605	.00325	.00468	.0141	
R				1.32	1.20	1.54	1.51	1.41	1.38	1.32	1.165	
Model temp., °C	cylinder		159.5	171.3	174	126.2	130.4	122.5	188.2	180	119.8	121.2
	end		152.0	165.5	169	169.9	123.8	103.7	180.8	173	113.7	114.8
	average			169.9	172.8	124.6	128.8	117.8	186.4	178.3	118.3	119.6
Wall temp. °C	cylind.r, °C		30.0	50.5	61.8	40	59[4]	-29	54	59	39	39
	end, °C		30.0	41.0	40.0	36	36	-38	36	45	35	34
Air temperature, °C				35.0	28.0	37	38	-38	40	50	37	36

Footnotes: 1. Heated drum
2. Model perpendicular to air flow
3. Calculated from Figure 10 and equation (4)
4. Brass rods

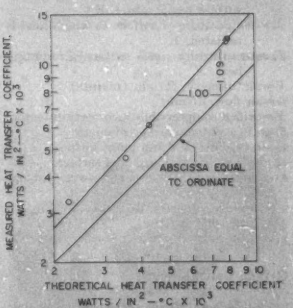

MEASURED HEAT TRANSFER COEFFICIENT, WATTS / IN 2 — °C X 10 3
THEORETICAL HEAT TRANSFER COEFFICIENT WATTS / IN 2 — °C X 10 3
1.09
1.00
ABSCISSA EQUAL TO ORDINATE

Fig. 10. Comparison of measured and theoretical forced convection heat transfer coefficients

the test model were greater than the theoretical value and were more sensitive to air flow changes than the theory indicates.

4. Equations are derived which can be used to calculate the convective and radiative heat transfer from the surface of the test model. The test model can, therefore, be used as a calibrated device to check the thermal environment in altitude chamber and air-borne installations.

Appendix

Nomenclature

a = surface area, in.^{2}
A = constant
B = constant
d = diameter of drum, ft
d_1 = diameter of orifice, ft
D = equivalent diameter of model $\left(D=\frac{a}{144\pi}=0.788 \text{ for test model}\right)$
g = gravitational constant, 4.17×10^8 ft/hr^2
h = heat transfer coefficient, watts/in.^{2}, K (degrees Kelvin)
k = thermal conductivity, watts/in., K (degrees Kelvin)
K = orifice coefficient
N_G = Grashof number
N_P = Prandtl number
N_R = Reynolds number
N_u = Nusselt number
p = pressure drop across orifice, inches of water
P = ambient pressure, centimeters mercury absolute
P_1 = static pressure upstream of orifice, centimeters mercury absolute
q = heat dissipation, watts
R = measure h/calculated h
t_c = average temperature of cylindrical surface, C (degrees centigrade)
t_m = average temperature of test model, C
t_u = average temperature of vertical surface, C
T = air temperature, K
T_f = air film temperature, equal to arithmetic average of air and model surface temperature, K
T_r = temperature of surface to which heat is radiated, K
T_s = temperature of surface radiating heat, K
V = air velocity, ft/min (minute)
w = air flow, lb/min
Y = orifice compressibility correction factor
β = orifice diameter/pipe diameter
ϵ_d = emissivity of drum inner surface
ϵ_m = emissivity of test model surface
θ = temperature difference between object being cooled and heat sink, C
μ = absolute viscosity, lb/ft-sec
ρ = air density, lb/ft^3
ρ_1 = air density upstream of orifice, lb/ft^3

Orifice Calculations

The purpose of this section of the paper is to convert the method for calculating air flow into a form easier to use.

Equation 175 of reference 4, converted to the units used in this paper, becomes

$$w=14.36KYd_1^2\sqrt{\rho_1 p} \tag{6}$$

For air

$$\rho=0.290\frac{P_1}{T} \tag{7}$$

For Reynolds number, equation 129, reference 4, converted to the proper units, becomes

$$N_R=\frac{4w}{\pi d_1\mu} \tag{8}$$

Table XVIII of reference 8 contains a tabulation of the viscosity of air as a function of temperature. For the pressure range under investigation, the viscosity is independent of pressure. For the temperature range being considered in this study (−55 to 250 C), the viscosity can be represented by the following equation:

$$\mu=\frac{4.67}{10^4}T^{0.736} \tag{9}$$

where μ is lb/ft-hr and T is in F (degrees Fahrenheit) absolute.

The power function approximation was used instead of the exact Sutherland equation (equations 5-6 of reference 8) because it results in a much simpler final equation. Converting to the units used in this paper, equation 9 becomes

$$\mu=\frac{1.987}{10^7}T^{0.736} \tag{10}$$

Substituting equations 6, 7, and 10 in equation 8 results in the equation

$$\frac{N_R}{K}=4.97\times10^7\frac{\sqrt{pP_1}\,d_1Y}{T^{1.236}} \tag{11}$$

Tables 6, 7, and 8 of reference 4 list K as a function of β and Reynolds number. For a given installation, constant β and pipe size, plot K as ordinate and $\frac{10{,}000}{(N_R/K)}$ as the abscissa. Using a straight line approximation, the constants for the following equation can be determined:

$$K=A+B\left[\frac{10{,}000}{N_R/K}\right] \tag{12}$$

This correlation method is a modification of the method illustrated in Fig. 26 of reference 4.

By substituting equations 11 and 12 in equation 6, an equation for air flow in the following form is obtained:

$$w=7.74YAd_1^2\sqrt{\frac{pP_1}{T}+\frac{1.557Bd_1T^{0.736}}{10^3}} \tag{13}$$

Assuming a uniform air velocity in the drum, the continuity equation becomes

$$V=264\frac{wT}{Pd^2} \tag{14}$$

For the system used in this experiment:

$d_1=1/3$ ft
$\beta=0.592$
A = 0.648 ft^2
B = 0.090 (this term can be neglected)
$d=2$ ft

Combining equations 13 and 14 and substituting the foregoing constants, the equation for flow velocity becomes

$$V=36.8\frac{Y}{P}\sqrt{pP_1T} \tag{15}$$

Figs. 72 and 73 of reference 4 contain plots of Y as a function of β and p/P.

Natural Convection

By combining equations 25-3 and 25-4 of reference 6, the results of experiments on the natural convection from horizontal cylinders and spheres can be correlated by the following equation:

$$N_u=0.5\sqrt[4]{N_GN_P}+0.4 \quad \text{For air} \tag{16}$$

$$N_P=0.72 \tag{17}$$

$$N_G=\frac{gD^3\rho^2\theta}{\mu^2T_f}, \text{ and} \tag{18}$$

$$N_u=\frac{hD}{k} \tag{19}$$

To maintain a constant set of units, the derivation will be made in a consistent set of English units and then converted into the units used in this paper. In this case,

$$\rho=0.522\frac{P}{T}, \text{ where } T \text{ is in F absolute} \tag{20}$$

Using the data contained in Table XVIII reference 8, and the method of analysis used to derive the viscosity equation, the thermal conductivity for air can be represented by the following equation:

$$k=\frac{1.399}{10^4}T^{0.743} \tag{21}$$

where k is in Btu/ft-hr F absolute.

Substituting equations 17, 18, 19, 20, 21 in equation (16) we get

$$h=\frac{0.308\sqrt{P}\sqrt[4]{\theta}}{\sqrt[4]{D}\,T^{0.370}}+\frac{0.56T^{0.743}}{10^4D} \tag{22}$$

where h is in Btu/ft^2 hr F absolute and θ is in F.

Converting units, equation 22 becomes

$$h=\frac{10.5\sqrt{P}\sqrt[4]{\theta}}{10^4\,T_f^{0.370}\sqrt[4]{D}}+\frac{0.00567}{10^4}\frac{T_f^{0.743}}{D} \tag{23}$$

Forced Convection

Reference 5 uses the following equation to represent the forced convection heat transfer between spheres and air:

$$\frac{hD}{k}=0.33(N_R)^{0.6} \tag{24}$$

From page 99 of reference 4,

$$N_R=\frac{D\frac{V}{60}\rho}{\mu} \tag{25}$$

Using the equations for ρ, k, and μ and combining terms

$$h=\frac{1.12}{10^4}\frac{(PV)^{0.6}}{T_f^{0.293}D^{0.4}} \tag{26}$$

Temperature Averaging

For this section of the paper, the temperature at a given thermocouple location, see

Fig. 9, will be represented by that number enclosed in a circle.

For the cylindrical surface

$$t_c = \frac{①}{4} + \frac{③}{8} + \frac{⑤}{4} + \frac{④}{4} + \frac{⑤}{8} \quad (27)$$

for the vertical surface

$$t_v = \frac{②}{4} + \frac{③}{2} + \frac{⑤}{4}, \text{ and} \quad (28)$$

for the average temperature of the entire model

$$t_m = \frac{t_v}{4} + \frac{3}{4} t_c \quad (29)$$

Heat Loss

To calculate total heat loss

$$q = ha\theta \quad (30)$$

References

1. Correlation of the Thermodynamic and Electrical Characteristics of Blast-Cooled Generators, Daniel Friedman. *AIEE Transactions*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 448–56.

2. High Altitude Aircraft Inverters, P. W. Franklin. *Ibid.*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 363–72.

3. Service Evaluation of Leland Electric Inverter, Type SE 20-2, 1500/1250 VA, R. W. Lindsley. *Report No. 2*, Naval Air Test Center, Patuxent River, Md., Sept. 28, 1956.

4. Fluid Meters, Their Theory and Application. *ASME Research Publication*, American Society of Mechanical Engineers, New York, N. Y., 1937.

5. Heat Transmission (book), William H. McAdams. McGraw-Hill Book Company, Inc., New York, N. Y., 1942.

6. Heat Transfer (book), Max Jacob. John Wiley & Sons, Inc., New York, N. Y., 1949.

7. Influence of Turbulence on Transfer of Heat from Cylinders, J. Koster, P. F. Maeder. *NACA Technical Note 4018*, National Advisory Committee for Aeronautics, Washington, D. C., Oct. 1957.

8. Heat Transfer (book), M. Fishenden, O. A. Saunders. Oxford University Press, London, England, 1950.

Co-ordination of Hydraulic Transmissions and Aircraft A-C Generators

S. C. CALDWELL
ASSOCIATE MEMBER AIEE

T. E. COPPINGER
NONMEMBER AIEE

R. T. SMITH
ASSOCIATE MEMBER AIEE

IT HAS BEEN demonstrated in the past that hydraulic variable-input-speed constant-output-speed transmissions can fail under certain fault conditions. However, occurrences in flight with class *A* and class *B* systems have not been, to the authors' knowledge, a major source of trouble. As shown in this paper, the possibility of drive failure is greatly increased in applications of class *C* systems. In fact, one such failure has reportedly occurred in flight. Hence it is essential to co-ordinate drive and electric system design to obtain the high electric system reliability required in modern aircraft. This paper presents a co-ordinated design on one system and introduces a general method for analysis which is sufficiently accurate to permit system design without extensive redesign of components after model testing has been completed.

System Description

The system considered herein is an 8,000-rpm 20-kva 0.75-pf (power factor) class *C* system. The generation unit is an integrated unit (in which the generator has only one bearing) and the generator is cooled by drive oil after it passes through the drive. The unit is designed for parallel operation and has provisions for both real and reactive load division as well as for an overrunning clutch of the sprag type.

A mechanical underspeed switch is incorporated on the drive output. Its contacts operate at about 380 cycles on increasing speed and 360 cycles on decreasing speed. An overspeed sensing device is incorporated in the drive output. This device shifts the drive into full underdrive when output speed exceeds that corresponding to 480 cycles. Static excitation controlled by a high-phase takeover regulator is used. The regulator maintains the average of the line-to-neutral voltages at rated value until voltage unbalance is such that one voltage is about 3 volts higher than average. For unbalances beyond this point, the regulator maintains the highest phase voltage near rated value regardless of what average values are unless excitation ceiling limits this voltage to less than rated. Fig. 1 shows a simplified diagram of the mechanics of the drive and Fig. 2 shows the integrated unit. The static exciter-regulator and protective panels are separate units which may be mounted in any convenient location in the airplane.

Torques Imposed on the Drive by the Electric System

Large electrical torques in an alternator result from three situations: initial torque peaks resulting from sudden application of fault; steady-state overload and fault torques; and synchronizing torques. The drive must be designed to withstand these torques without failure and, in some applications, speed must not sag below a specified value.

Drive operating pressure is directly proportional to torque transmitted by the oil. Since this pressure determines the magnitude of peak working stresses, it is necessary to establish the size, duration, and frequency of this transmitted torque and

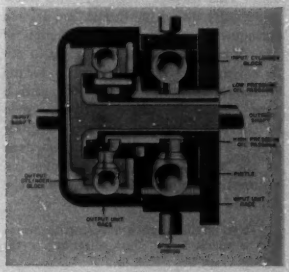


Fig. 1. Schematic diagram, constant-speed drive

Fig. 2. Integrated drive and alternator

Paper 58-848, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 13, 1958; made available for printing April 10, 1958.

S. C. Caldwell and T. E. Coppinger are with the General Electric Company, Schenectady, N. Y.; R. T. Smith is with the University of Texas, Austin, Tex.

temperatures at which they occur to establish drive capabilities in meeting these requirements.

The authors have sought to obtain means for determining electrical torques imposed on the system and the resulting drive torques so that a co-ordinated design may be obtained.

It is expected that the co-ordination methods derived will result in better system performance and reliability consistent with light weight than has been obtained in the past.

Initial Fault Torques

Sudden application of a fault causes very-high short-duration torques to appear on the rotor. Appendix I shows the method of calculating these for balanced faults (including effects of excitation forcing). The curves of Figs. 3(A) and 3(B) show these torques for a typical aircraft machine for two different values of fault resistance during the first 10 milliseconds. Continuation of these curves shows only a relatively slow rise to steady-state values from the value of 10 milliseconds resulting from excitation build-up to steady-state values. This is shown in Fig. 4.

Hydraulic drives, without governor action, droop with load as a result of leakage. This leakage, together with alternator rotor and drive inertia and system springiness as shown in the diagram of Fig. 5, gives the drive the ability to act as a hydraulic shock absorber for torque transients faster than governor reaction time. Thus, the transient torques seen by the drive may be expected to differ greatly from the electrical torques imposed on the alternator rotor. The following assumptions are made to simplify the study of these torques.

1. In jet engine applications, engine inertia (J_e) is much larger than alternator rotor inertia (J_m) and drive output inertia (J_d). Hence engine inertia will be considered to be infinite.

2. Speed controls on hydraulic drives produce response in the range of 0.1 second to 1.0 second. Therefore, it is assumed that in the first 10 milliseconds, no governor action is obtained.

3. The frame and stator assembly of the integrated unit is rigid.

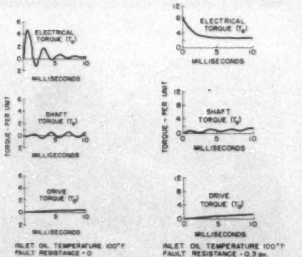


Fig. 3. Initial system torques with 3-phase faults

With these assumptions, the analog computer representation shown in Fig. 6 may be derived easily from the schematic of Fig. 5. In Fig. 5,

$J_e = \infty$ = main engine inertia

$J_d = 0.012$ pound-inch-second^2 = drive output end inertia

$J_a = 0.186$ pound-inch-second^2 = alternator rotor inertia

$K_s = 5.8 \times 10^4$ pound-inch = spring constant of drive output and alternator shafting

$K_o = 3.2 \times 10^2$ pound-inch = effective spring constant of drive oil

$13 > R > 3.8$ pound-inch-second = damping coefficient of drive (a function of drive leakage)

It is, of course, true that the damping factor (R) would increase sizeably at lower oil temperatures and would approach infinity at −65 F (degrees Fahrenheit). Hence the shock-absorbing capability of the drive would approach zero. However in an ambient of −65 F, engine start-up will require an estimated 30 seconds. In this period, drive oil temperature will rise to about 300 F before the output is up to speed since oil in the lines to the cooler cannot circulate rapidly enough initially to cool the drive. Hence oil is recirculated within the drive itself and heats rapidly. As time passes, more and more oil circulates through the cooler until drive oil temperature stabilizes at about 100 F with −65 F cooling conditions. Thus, the minimum shock-absorbing power of the drive is with inlet oil at 100 F for any operating condition. At higher oil temperatures, the shock-absorbing capabilities of the drive are greater since the oil thins and leakage is greater. It is seen, however, that alternator winding temperature, for a very short time during this start-up period, may be at essentially −65 F even though drive oil is hot. In summary then, operating oil temperature at the drive inlet varies from 100 F up to the rated maximum of 300 F.

Results of the computer study are shown in Fig. 3 for drive torque build-up during the initial period and Fig. 4 shows drive torque build-up over a longer period.

The curves shown neglect accelerating torque caused by governor action in recovering speed after the initial dip. However, at these large torques [above 2.0 pu (per unit)], drive response is so much slower than normal that speed build-up takes place over a longer period and acceleration torques during speed recovery are not expected to exceed about 0.2 pu.

Examination of these runs shows that the shock-absorbing action of the drive integrates out the very large initial torque peaks. In fact, during the period in which these very-high short-duration electrical torques exist, drive torque is negligible and the contribution of these peaks to drive torque at any time is small.

It is seen, however, that a lightly damped circuit exists in the system. This circuit consists of the alternator rotor inertia, the drive output end inertia, and the connecting shafting. Since very sizeable fundamental frequency electrical torques can exist in the initial period and damping in this subsystem is small, it is necessary to design this portion to resonate as far from 400 cycles as is consistent with shaft size to carry the torques. Otherwise, shaft breakage may ensue even though drive working pressure may not be exceeded. With unbalanced faults, both fundamental and second harmonic electrical torques are very large in the initial period and higher harmonics are also present. Hence design must be such that this portion of the circuit does not resonate at fundamental or higher harmonic frequencies.

These runs have all been based on 3-phase fault conditions. In the unbalanced fault cases, initial electrical torques have the same characteristics of

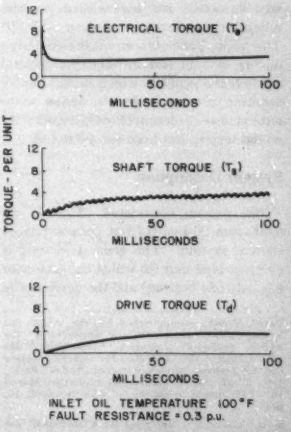


Fig. 4. System torques with 3-phase faults

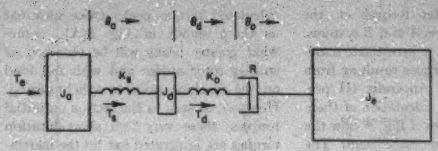


Fig. 5. Schematic diagram of rotating components

very-high short-duration peaks coupled with large first and second harmonic components. Again, these transients die out rapidly and torques seen by the drive as a result of these short-duration conditions are very small compared to steady-state values.

It is concluded, then, that drive design can be based on steady-state values and that initial electrical torques, even though large, need not be considered.

However, even though the rotating part of the system is, essentially unaffected by these initial torques, the reaction torques on the stator and frame of the integrated unit must be considered since the frame is a fairly stiff member and no shock-absorbing mechanism is available for cushioning the impact.

With rigid fastenings between the stator of the alternator and the frame and a rigid frame for the unit, these peak torques would be felt down the length of the frame into the engine pad. Frames and fastenings are not completely rigid though and these peak electrical torques are modified by the effect of the inertia-spring system equivalent of the stator, stator fastenings, frame, and engine pad. Since the frame is of large diameter it is capable of sustaining very large torques. It also has an appreciable distributed inertia. Thus it is likely that the greatest stresses will be imposed on the fastenings between stator and frame in the alternator.

It would appear that machine designers have taken care of this adequately in the past in that there seems to be no case in which stator fastening breakage has been traced to a fault condition. In extensive laboratory testing, including a large number of all types of faults on engineering samples of the system described later, there has been no evidence of failure or of incipient failure of any of the fastenings.

If no drive is used and the alternator is coupled directly to the engine pad through stiff couplings, the shock-absorber action of the hydraulic drive is not present and peak shaft torques may become quite high.

Steady-State Fault Torques

Although it has been shown that the very-high short-duration initial torques resulting from sudden application of faults do not cause large drive torques, steady-state torques created by machine and fault losses can be very sizeable.

With unbalanced faults in steady-state, there are pulsating harmonic torque components. However, the rotating system is a very effective filter for harmonic torques for the same reasons that enable the system to absorb the large initial torques without any appreciable torque being introduced into the drive. A Bode plot of the rotating system shows that only 0.55% of second harmonic electrical torque is seen by the drive and, of course, as frequency goes higher, even less is introduced. Since subtransient saliency is small in most machines, second harmonic torques are not large and, therefore, 0.55% of second harmonic torque is a negligible factor in evaluating steady-state torques. Similarly, higher harmonic components may be neglected. It is concluded that only d-c components of torque need be considered. Again it should be noted that the lightly damped resonant circuit of drive output inertia, shafting to alternator, and alternator should not resonate at any of the harmonic frequencies if excessive shaft torques are to be avoided.

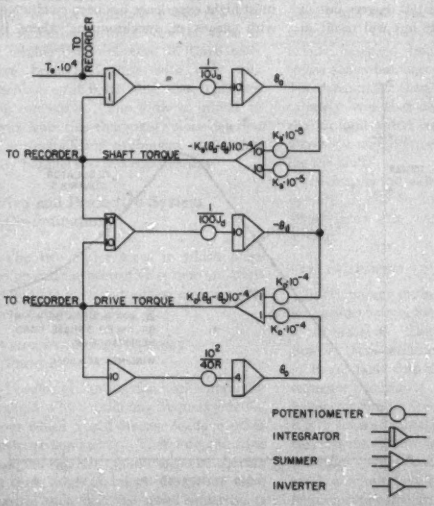


Fig. 6. Computer representation of rotating components of the system

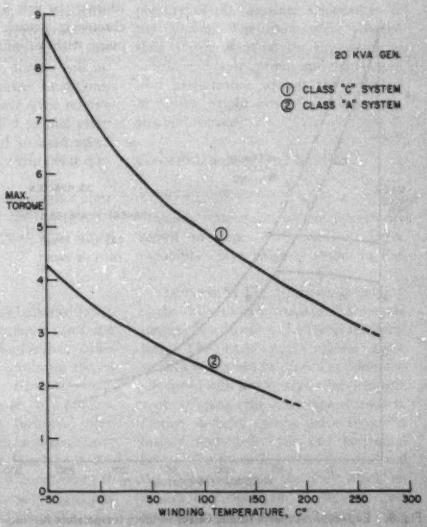


Fig. 7. Maximum torque versus winding temperature for 3-phase fault

Appendix II presents a simple method by which maximum steady-state torques may be determined for balanced 3-phase faults. It may be noted that maxima are obtained at rather sizeable values of balanced 3-phase resistances in the path of the fault current. Curves of maximum balanced fault power as a function of rotor temperature are shown in Fig. 7. It is seen that maximum torques with a class *C* system are sizably larger than is the case with a class *A* system. This results from the much greater excitation capabilities of the class *C* system.

The class *C* system must produce 3-pu 3-phase fault current with the rotor at 250 C (degrees centigrade). Since most excitation systems are rather stiff sources at ceiling conditions, the reduced rotor resistance at −50 C rotor temperature results in about twice the field current and 3-phase fault current will be about 6 pu. In effect, field current ceiling is about twice what it is at high rotor temperature. A class *A* system can qualify by delivering 3-pu 3-phase fault current at room temperature. Hence the added excitation available at low rotor temperatures is not great. Fault current capability at −55 C will be only about 4 pu. On the other hand, fault current at high rotor temperature will be substantially less than 3 pu.

In summary the class *C* system gives the electric system the capability of imposing sizably greater torques on the drive than do the class *A* and *B* systems.

It is necessary also to investigate maximum steady-state torques resulting from unbalanced faults. Appendix III presents the method for calculation of these torques and the curves of Fig. 8 show the results for the system considered. The torque curves of Fig. 8 show sharp breaks. These result from the fact that, at high rotor temperatures, excitation ceiling limits current and torque but, at lower temperatures, more excitation is available so torque increases until terminal voltages rise to the point where regulation starts reducing excitation. As a result, the maximum torques seen by the drive are determined by regulator action; thus, highest phase sensing has reduced torque maxima sizeably at lower temperatures.

Synchronizing Torque

Past practice has been to test a system for its ability to withstand random synchronizing a few times but not to require that it be able to withstand random synchronizing as a normal operational procedure. Synchronizing lights and automatic synchronizing devices have been used to prevent random synchronizing in normal operation. However, the shock-absorbing capabilities of the drive studied led the authors into the following investigation of the system reaction to random paralleling.

It can be shown that initial torques (during the first 5 milliseconds) on synchronizing a machine 180 degrees out of phase with an infinite bus will result in calculated torque peaks twice as great as those shown in Fig. 3(A). Somewhat greater peaks will be obtained at smaller rotor angles and with full load on the machine prior to synchronizing. However, as seen in the section on initial torques, these very-high short-duration torques are integrated out by the inertia-dashpot-spring rotating system and only very small torques are actually presented to the drive during this initial period. The 10- to 16-pu electrical torque peaks calculated by the means used widely in utilities systems have little meaning as far as the drive is concerned.

During the subsequent period when the rotor is moving into synchronism and excitation is increasing due to the action of the regulator and load division signals, torque analyses are very complex since the dynamics of the mechanical system, the drive and its controls, and the excitation system are all involved. No analysis has been attempted herein. Examination of oscillograms from other systems together with very rough estimates of the effects of such transients on this system indicate maximum torque presented to this drive during synchronizing should not exceed about 2 pu.

In considering torques imposed on the stator, stator fastenings, frame, and engine pad, the initial torques must be considered for the same reasons as those listed under fault torques. Since these initial torque maxima are of the order of twice those encountered in faults and since most flight experience has been conducted with means of synchronizing, there is

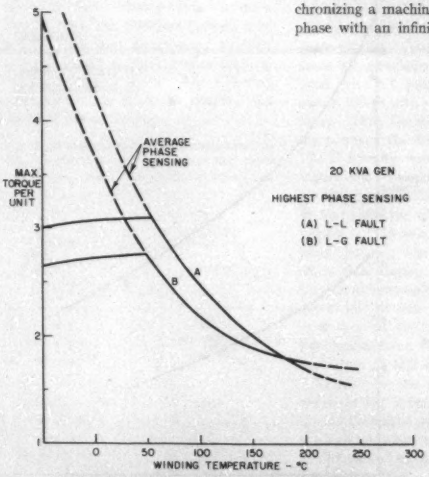


Fig. 8. Expected maximum torque versus winding temperature for line-to-ground and line-to-line faults

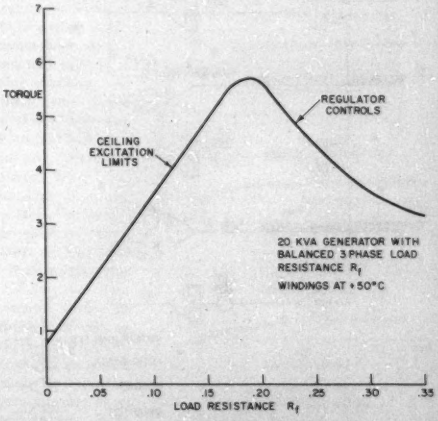


Fig. 9. Torque versus load resistance

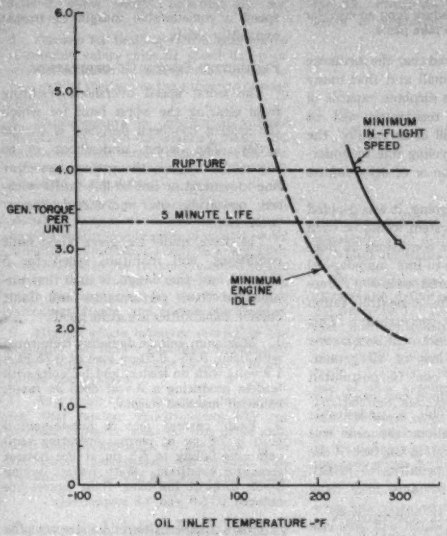


Fig. 10. Drive torque capability curves

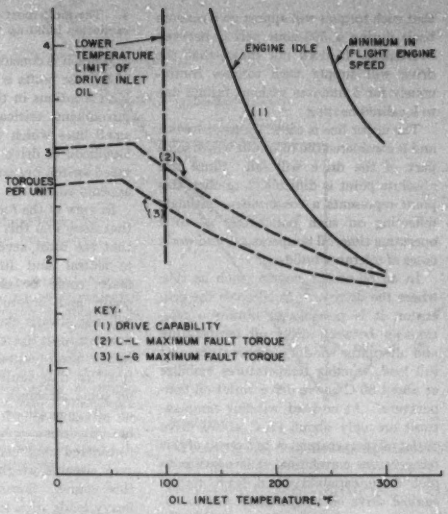


Fig. 11. Torque co-ordination curves

little operating experience to use as a measure of stator fastener, frame, and engine pad ability to withstand initial synchronizing torque as a normal operational procedure. In the laboratory, random synchronizing has been used thousands of times without ill effects. The pad mounting has not been the same as an engine pad and most of these tests have been performed with only two machines so it is felt that these tests are not conclusive. The authors intend to investigate this thoroughly since deletion of autoparalleling equipment would simplify parallel systems considerably.

Drive and Protective System Co-ordination

The two major areas in which drive design and the protective system are interrelated are over-and-underfrequency protection and fault clearing.

Over-and-Underfrequency Protection

Levels of protection are normally arrived at by considering frequency deviations which would damage loads or other system components. Therefore, the drive designer has but limited scope in arriving at these levels and must design the drive control such that the speed undershoots and overshoots corresponding to application and removal of the worst fault or overload torques considered applicable to the system does not cause the protection to trip the machine.

These levels of protection have been established as 360 cycles for underfrequency and 480 cycles for overfrequency protection. On application of a fault torque, as speed droops, torque will increase somewhat under many conditions. The drive must, then, have some reserve capacity over that needed to just carry the fault at rated speed to hold speed above the underfrequency trip point during the fault.

On clearing the fault, the governor must operate rapidly to hold the speed overshoot below the overspeed trip device setting.

Fault Clearing

Fault current during unbalanced faults is dependent upon the sequence constants of the machine. The trend is that reduction in these sequence constants results in larger fault current. However, small sequence constants tend to produce larger maximum fault torques. Since it is of some advantage to have large fault currents to obtain fast clearing of distribution system faults, it is necessary to reach a compromise between weight and fault current capability of the unit.

This is complicated by the fact that sequence constants of the machine also determine voltage unbalance with unbalanced loading. With 0.75-pf unbalanced loading, voltage unbalance is reduced by reducing the sum of the components of the sequence impedances of the system. The compromise selected thus is dependent on the evaluation of unbalanced fault current and unbalanced load performance against the penalties in system weight involved in improving this performance.

Co-ordination of One System

Torque Co-ordination

Steady-state fault torque curves are shown in Figs. 7, 8, and 9. Drive capability curve shape is shown in Fig. 10.

The curve in Fig. 10 shows the ability of the drive to carry certain steady-state torques as a function of oil-inlet temperature. The basic curve shows those torques which can be carried at minimum idle speed with just enough reserve to prevent tripping the underspeed switch. During in-flight speeds, drive torque is mainly limited by the two horizontal lines. The lower line has been designed the "5-minute drive life line" which indicates that operation at these torques seriously reduces over-all drive life. The 5-minute rating covers the total period

that such torques will appear on a random basis during a 500-hour period between overhauls. It does not mean that the drive will supply such torques continuously for 5 minutes without failure due to localized heating.

The upper line is called the rupture line and is considered the torque at which some part of the drive will fail. Since this absolute point is difficult to predict, this point represents a conservative estimate reflecting on such parameters as total operating time, oil temperature, and worst cases of tolerance build-up.

In an oil-cooled system, such as this, where the drive output oil cools the generator, it is possible to obtain a relationship between drive oil temperature and alternator winding temperature. At full load, winding temperatures stabilize at about 50 C above drive outlet oil temperature. At no load, winding temperatures are only about 12 C above drive outlet oil temperature. As a result of this temperature correlation, it is possible to plot drive capability and fault torque against drive oil temperature and to obtain torque co-ordination easily. Such a curve is shown in Fig. 11. It is seen in this figure that use of drive inlet oil for alternator cooling would shift fault curves to the right since the alternator would then be cooler than the drive. Hence, for the same electrical performance, use of drive outlet oil for alternator cooling saves weight in the drive at the expense of a comparatively small increase in excitation system weight.

Examination of Figs. 7 and 11 shows that drive torque capability must be doubled to carry maximum 3-phase fault torques. Since this adds 30% to drive weight, for the same input speed and temperature range, and engine accessory gear train and pad weights may increase too, the 3-phase fault case will be considered carefully to determine if its importance justifies the addition of, perhaps, 30 pounds' weight to the drive and main engine. Of course, in cases where the engine gearing and pad are already capable of carrying these torques at minimum drive input speed, additional weight will be involved only in the drive to the extent of about 12 pounds.

The requirements to produce these maximum 3-phase fault torques are

1. Fault resistance must be fairly symmetrical.

2. Fault resistance must lie in the range of 0.1 pu to 0.4 pu.

3. Inductance in the lines between the machine and the fault must be negligible or the torque will be reduced significantly since machine saturation is the limiting factor.

4. The fault must endure long enough for excitation build-up to take place.

When it is considered that the incidence of 3-phase faults is small and that many fault locations in the airplane capable of introducing critical resistances will be small lines which will clear rapidly, the likelihood of drive cycling due to underspeed switch operation or of drive failure appears small.

In view of the foregoing, it was decided that design on this system would be such that the most severe conditions of line-to-neutral and line-to-line unbalanced faults could be cleared safely. Three-phase and line-to-line-to-neutral faults will not always clear successfully. This does not meet the requirement that no one failure shall cause loss of all generation in that faults can be postulated on some airplanes which can cause loss of all generation. This is undesirable but, on this application, the risk was considered small enough in the face of the very sizeable weight penalties to justify this course. Some systems may have heavy loads with low negative and zero-sequence reactances, which will give rise to fault torques greatly in excess of those with static loads. A partial answer to this may be a torque-limiting device which will limit excitation automatically at any temperature to those values producing the maximum torque the drive can carry. If our present thinking is correct, this can be done without reducing 3-phase fault current seriously below 3-pu for faults in most parts of the airplane.

As seen in Fig. 11, the drive and electric system have been co-ordinated on the basis of the worst possible line-to-neutral and line-to-line fault conditions as follows:

1. The machine has been running at no load and rated load at rated power factor has been switched on just prior to the fault.

2. For every oil temperature, the fault resistance is such as to produce the maximum torque the alternator can generate. (In general, this critical fault resistance varies with machine temperature.)

3. The impedance encountered by the fault current is pure resistance.

4. The drive has had 500 hours' wear on it. (Wear has been checked by disassembly and measurement of a drive with 500 hours' operation.)

5. The fault is not cleared before steady-state torques are felt by the drive.

With this, it is felt that the system is capable of handling all possible cases of the fault types considered. It may be noted that co-ordination is obtained at engine idle for all oil temperature conditions; under minimum in-flight engine speed a considerable margin of torque capability exists.

Protective System Co-ordination

The worst speed overshoot resulting from clearing the worst fault for which the system has been designed is to 460 cycles. The worst undershoot is to 370 cycles. Thus, the system can clear line-to-neutral or line-to-line faults without operating over-and-underfrequency protection.

The drive, under the given worst fault conditions, will maintain speed for 5 seconds and the design is such that unbalanced load performance and fault current capabilities are quite good:

1. Maximum voltage deviation from rated with 0.35-, 0.35-, 0.50-pu load at 0.75 pf is 1.2 volts with no feeders and 1.7 volts with feeders producing a 5-volt drop at rated, balanced machine output.

2. Fault current into a line-to-neutral fault is 7.8 pu at normal operating temperatures falling to 6.3 pu at the hottest generator condition. With feeders having 5-volt drop at full load, these figures are reduced to 5.8 and 4.8 respectively.

This would seem to reflect a reasonable balance between drive weight, unbalanced loading performance, and fault current capabilities.

System Stability Co-ordination

A full-scale, detailed stability analysis of the whole system including the mechanical and hydraulic characteristics, the drive control, the excitation system, the alternator, and all the paralleling controls has been made. The study (on an analog computer) showed that the important parameters are in drive speed control and load division rather than in excitation control characteristics. These latter characteristics can be varied widely without affecting speed or real load division stability (for example, regulator gain can be varied over a range of 2 to 1 without producing instability).

Engineering samples in test have produced no evidence of periodic oscillations and transients appear well damped over the whole range of operation.

Conclusions

1. An integrated drive and alternator unit which meets the needs of the application has been designed.

2. Drive and engine pads (particularly in larger system ratings) cannot be designed to withstand maximum steady-state torques that a class C alternator can develop without serious weight penalties. If 3-phase and line-to-line-to-neutral faults (under the worst possible conditions) and loads of unequal sequence impedances are not con-

sidered, these weight penalties are not incurred.

3. Devices for limiting torque maxima to reasonable values without speed droop or serious reduction in fault current appear to be a partial solution to the problems of conclusion 2.

4. Improvements in unbalanced voltage performance (with unbalanced load) and unbalanced fault current capabilities usually increase maximum unbalanced fault torques presented to the drive.

5. The very-high short-duration peak electrical torques developed by the generator on sudden application of a fault or by random synchronizing result in a negligible contribution to the torque presented to the drive. These initial torques, however, must be considered in alternator stator fastenings, frame, and engine pad designs.

6. Highest phase takeover sensing in the regulator can reduce maximum unbalanced fault torques by as much as 2 to 1 in air-cooled applications.

7. As a result of conclusion 5, the authors feel that random synchronizing will not produce excessive wear on this drive and, with suitable attention to generator frame and stator construction, will become both feasible and desirable in many applications.

Appendix I. Initial Fault Torques

From a book on synchronous machinery by Concordia,[1] the following general expression for electrical torque on a 3-phase fault may be derived easily.

$$T = e^2\left[\frac{F}{X_d''}e^{-t/T_a'}\sin\omega t + \frac{1}{2}\left(\frac{R_d}{X_d''^2}+\frac{R_q}{X_q''^2}\right)e^{-2t/T_a'} + \left(\frac{F}{X_d''}+I_d\right)^2 R_a' + \frac{I_d^2 R_a'^2}{X_q''^2} + \frac{1}{2}\left(\frac{1}{X_q''}-\frac{1}{X_d''}\right)e^{-2t/T_a'}\sin 2\omega t\right]$$

where

$$F = \left(1-\frac{X_d''}{X_d'}\right)e^{-t/T_d''} + \left(\frac{X_d''}{X_d}\right)e^{-t/T'd2} + \frac{X_d''}{X_d}$$

$$T_a' = \frac{R_a}{R_a'}T_a$$

$$R_a' = R_a + R_x$$

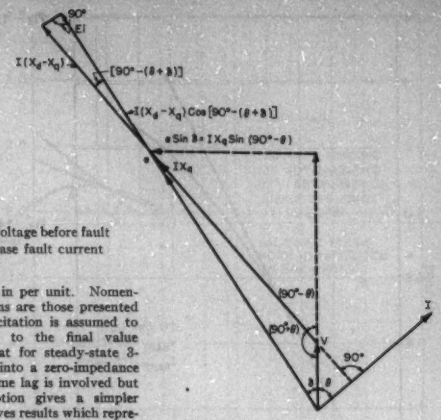
$$R_q = \text{Real}\left\{\frac{-X_{aq}(X_{kq}-jR_{kq})}{R_{kq}+j(X_{kq}+X_{aq})}\right\}$$

$$R_d = \text{Real}\left\{\frac{X_{ad}^2 R_{kd}+R_f+j(X_f+X_{kd})}{R_{kd}R_f-X_{ad}(X_f+X_{ad})-X_{kd}X_f+j[R_f(X_{ad}+X_{kd})+R_{kd}(X_{ad}+X_f)]}\right\}$$

$$I_d = \frac{X_q(\alpha X_q-1)(1-e^{-t/T'd2})}{X_d X_q + R_a'^2}$$

$$T' \quad e' \frac{R_a'^2+X_d'X_q}{R_a'^2+X_dX_q}$$

Fig. 12. Phasor diagram of alternator



e = no-load terminal voltage before fault
α = steady-state 3-phase fault current
R_x = fault resistance

All quantities are in per unit. Nomenclature and definitions are those presented by reference 1. Excitation is assumed to rise instantaneously to the final value corresponding to that for steady-state 3-phase fault current into a zero-impedance fault. Actually a time lag is involved but use of this assumption gives a simpler analysis and, also, gives results which represent the worst case. If it is desired to factor in excitation delay, it may be inserted in the defining equation for I_d by one of the methods given in the reference.

Appendix. II. Steady-State Torque Calculations for 3-Phase Faults

Experience has shown that the steady-state torques imposed upon drives by the electric system may be very high, to the point of being destructive. These torques arise when abnormally small impedances are applied to the system and the voltage regulator attempts to maintain normal terminal voltage. That very high excitation currents may exist is in large measure caused by the large capabilities of the exciter at an elevated rotor winding temperature and the fact that under certain conditions the rotor winding may be at a very low temperature, perhaps −55C.

Under balanced load conditions the torques may be calculated quite accurately since the symmetry of the system greatly simplifies the mathematical representatives of the physical phenomena. An analysis may be carried out as follows.

Assumptions

1. Unless otherwise noted, the voltage regulator maintains rated terminal voltage until ceiling excitation is reached.

2. Machine speed is constant.

3. Saturation is assumed to be a function of air-gap voltage only.

4. Saturation effects may be accounted for by an effective decrease in excitation.

5. Friction, windage, hysteresis and eddy current losses, and exciter power requirements have not been considered. Together, they may constitute a measurable increase over the delivered power.

6. Armature resistance voltage drop is negligibly small with respect to machine reactance drop.

Analysis

From Fig. 12, the following equations are obtained easily:

$$e^2 = V^2 + I^2X_q^2 - 2VIX_q\cos(90°+\theta) \quad (1)$$

$$V = ZI \quad (2)$$

$$e\sin\delta = IX_q\sin(90°-\theta) \quad (3)$$

$$E_f = e + I(X_d - X_q)\cos(90°-\theta-\delta) \quad (4)$$

$$P = \frac{E_fV}{X_d}\sin\delta + \frac{V^2}{2X_dX_q}(X_d-X_q)\sin 2\delta \quad (5)$$

where

V = terminal voltage
E_f = excitation voltage
e = vector sum of V and IX_q
I = magnitude of load current
Z = magnitude of load impedance
X_d = direct axis reactance
X_q = quadrature axis reactance
θ = phase angle of load impedance
δ = power angle of the machine
P = power output of the machine

Whenever operating conditions are such that voltage regulator action is not effective in maintaining steady-state terminal voltage, it is desirable to express power and torque in terms of the machine constants, the excitation, and the load impedance.

From equations 1, 2, and 3,

$$e = I\sqrt{Z^2+X_q^2+2ZX_q\sin\theta} \quad (6)$$

$$\sin\delta = \frac{X_q\cos\theta}{\sqrt{Z^2+X_q^2+2ZX_q\sin\theta}} \quad (7)$$

$$\cos\delta = \frac{Z+X_q\sin\theta}{\sqrt{Z^2+X_q^2+2ZX_q\sin\theta}} \quad (8)$$

Combining these results with equation 4,

$$\frac{E_f}{I} = Z_e = \frac{Z^2+Z(X_d+X_q)\sin\theta+X_dX_q}{\sqrt{Z^2+X_q^2+2ZX_q\sin\theta}} \quad (9)$$

where

Z_e = total effective impedance seen by the excitation voltage

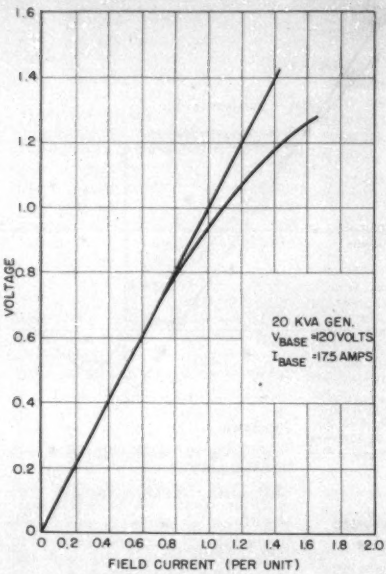


Fig. 13 (left). Alternator saturation curve

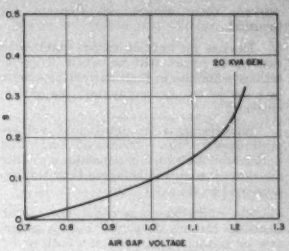


Fig. 14 (right). Saturation function for alternator

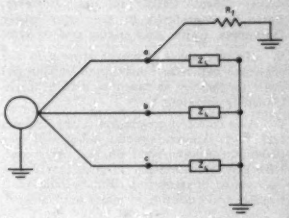


Fig. 15 (right). Schematic diagram of fault condition

Examination of equation 5 shows that all quantities may now be expressed in the desired terms and, after considerable algebraic manipulation, the following expression for power is obtained:

$$P=\frac{E_t^2 Z\cos\theta(X_q^2+Z^2+2X_qZ\sin\theta)}{|Z^2+Z(X_d+X_q)\sin\theta+X_dX_q|^2} \quad (10)$$

In Fig. 12, it is seen also that

$$E_a^2=V^2+I^2X_a^2-2VIX_a\cos(90°+\theta) \quad (11)$$

which, when combined with equation 2, reduces to

$$E_a=I\sqrt{Z^2+X_a^2+2ZX_a\sin\theta}$$

where

E_a = air-gap voltage
X_a = leakage reactance of armature

From equation 9,

$$I=\frac{E_t}{Z_s}$$

which combined with equation 11 gives

$$\frac{E_a}{E_t}=M=\frac{\sqrt{(Z^2+X_a^2+2ZX_a\sin\theta)\times(Z^2+X_q^2+2ZX_q\sin\theta)}}{Z^2+Z(X_d+X_q)\sin\theta+X_dX_q} \quad (12)$$

where

M = ratio of air-gap voltage to excitation voltage as defined by equation 12

Since it is assumed that saturation effects may be accounted for by subtraction of the difference in excitation for a given air-gap voltage between the air-gap line and the saturation curve of Fig. 13, this difference may be plotted against air-gap voltage as shown in Fig. 14 and an approximate equation relating the saturation and air-gap voltage may be deduced. One such equation would be

$$S=K(E_a-a)^2 \text{ (all quantities in per unit)} \quad (13)$$

for

$$E_a\geq a$$

with

K = a constant

For this machine, over the range of interest, $K=1.1$, $a=0.7$.

Then

$E_t=E_{to}-S$, and $E_a=ME_t$

and, combining with equations 12 and 13

$$M^2KE_t^2+E_t(1-2aMK)-E_{to}+a^2K=0 \quad (14)$$

where

E_{to} = actual excitation voltage in per unit
E_t = effective excitation voltage (corrected for saturation)

This quadratic equation may be solved and, when the negative root is discarded,

$$E_t=\frac{(2aKM-1)+\sqrt{4E_{to}KM^2-4aKM+1}}{2KM^2} \quad (15)$$

This value of E_t may then be used in the power equation 10 to obtain the power or torque for a given excitation with approximate compensation for saturation effects. It should be emphasized that equations 13, 14, and 15 must be used only when the air-gap voltage is equal to or greater than a per unit.

At loads too small to require maximum excitation, the voltage regulator will hold the terminal voltage to its rated value and the power equation becomes simply

$$P=\frac{V^2}{Z}\cos\theta$$

or, in per unit,

$$P=\frac{\cos\theta}{Z} \quad (16)$$

The power (and torque) curves may now be derived by the use of equations 10, 15, and 16. Starting from zero load impedance, equation 10 may be used until the air-gap voltage exceeds a per unit at which point equation 10 will be modified by a changed excitation voltage (E_t) derived from equation 15. From this point up to the load impedance where maximum excitation will produce rated terminal voltage, this modified equation 15 will hold; from the point where terminal voltage becomes rated on out to higher impedances, equation 16 will hold. The curve derived by this method is shown in Fig. 9.

Appendix III. Steady-State Torque Calculations for Unbalanced Faults

In order to calculate the steady-state unidirectional torques imposed upon the drive while an unbalanced fault exists on

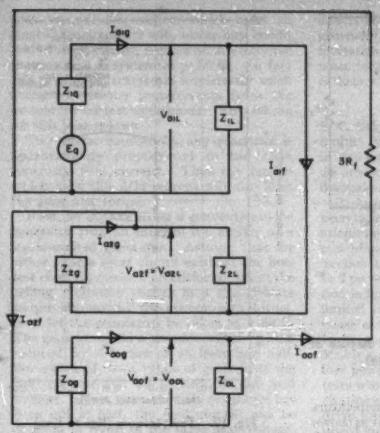


Fig. 16. Symmetrical component diagram of fault condition (line-to-ground)

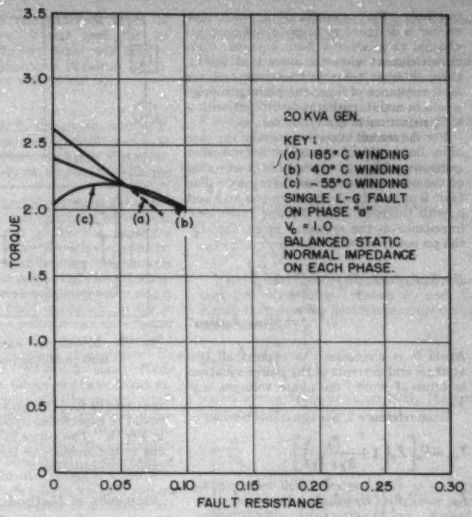


Fig. 17 (right). Fault torques (line-to-ground)

the electric system, it is advantageous to use the method of symmetrical components. By means of symmetrical components it is possible to compute the losses in the various portions of the electric system using readily available constants of the generator and lines.

The equations relating fault currents and fault voltages and their symmetrical components, for various types of faults, are presented in a number of standard references. Properly interpreted, these equations may be applied to the problem at hand. As examples of the method, the cases of a line-to-ground and a line-to-line fault will be carried out here in detail.

The system to be analyzed is sketched in Fig. 15. As shown, a symmetrical line-to-ground fault through a fault resistance R_f appears on phase a of a generator. The generator is also loaded by the rated 3-phase static impedance of 1.0 pu and 0.75 pf.

The following assumptions will be made:

1. Machine speed is constant.
2. The machine is fully symmetrical.
3. Friction, windage, hysteresis, and eddy losses, as well as excitation losses, have not been included.
4. Losses due to high harmonics are lumped into an efficiency factor.
5. Fault impedance is purely resistive.
6. Balanced, static, rated loads are at the generator terminals.

The sequence networks and interconnections are shown in Fig. 16. In Fig. 16,

$$Z_{1g} = R_{a1} + jX_g$$

$$Z_{1L} = \frac{R_{1L} jX_{1L}}{R_{1L} + jX_{1L}}$$

$$Z_{2L} = \frac{R_{2L} jX_{2L}}{R_{2L} + jX_{2L}}$$

$$Z_{0L} = \frac{R_{0L} jX_{0L}}{R_{0L} + jX_{0L}}$$

$$Z_{2g} = R_{a2} + jX_{2g}$$

$$Z_{0g} = R_{a0} + jX_{0g}$$

Then the total pu torque can be expressed as

$$P = R_{a1} I_{a1g}^2 + R_{a2} I_{a2g}^2 + R_{a0} I_{a0g}^2 + (R_{a2} - R_{a1}) I_{a2g}^2 + \frac{V_{a1L}^2}{R_{1L}} + \frac{V_{a2L}^2}{R_{2L}} + \frac{V_{a0L}^2}{R_{0L}} + \frac{R_f I_f^2}{3}$$

where I_f stands for the magnitude of the current through the fault resistance, and the second term in I_{a1g}^2 represents shaft input for negative sequence effects.

If it is assumed that speed changes are negligible, then the pu torque is also equal to the pu power, so that the expressions can be used interchangeably.

In order to calculate the power using equation 1, it is necessary to develop expressions for the various currents and voltages. Usually the machine ceiling excitation is a known quantity and must somehow be related to the foregoing currents and voltages. This process is complicated by the nonlinearity of the machine since saturation may exist. Therefore, it may be simpler to, (1) express the power equation quantities in terms of an assumed terminal voltage, say $\bar{V}_c$; (2) calculate the excitation for a series of values of $\bar{V}_c$; and (3) compute the power for the value of V_c which corresponds to the known excitation.

Proceeding with this idea, assume that phase c voltage is $1.0/\underline{0°}$; then, from reference 2,

$$I_{a1f} = [3aR_f + j\sqrt{3}(Z_2 - a^2 Z_0)]^{-1}$$

$$V_{a1f} = I_{a1f}(Z_0 + Z_2 + 3R_f)$$

$$\bar{V}_{a2f} = -I_{a1f} Z_2$$

$$\bar{V}_{a0f} = -I_{a1f} Z_0$$

$$I_f = 3I_{a1f}$$

$$I_{a1f} = I_{a2f} = I_{a0f}$$

Now from Fig. 16,

$$I_{a1g} = I_{a1f} + \bar{V}_{a1L}/Z_{1L}$$

$$= I_{a1f}\left[1 + \frac{Z_0 + Z_2 + 3R_f}{Z_{1L}}\right]$$

$$I_{a2g} = I_{a2f}' \bar{V}_{a2f}/Z_{2L}$$

$$= I_{a1f}[1 - Z_2/Z_{2L}]$$

$$I_{a0g} = I_{a0f} + \bar{V}_{a0f}/Z_{0L}$$

$$= I_{a1f}[1 - Z_0/Z_{0L}]$$

All the equations necessary to solve for the torque are now present. The quantities Z_0 and Z_2 are the sequence impedances as viewed from the point of fault. They are seen to be the parallel combinations of the generator and load impedances.

The results of a calculation for a 20-kva machine are shown in Fig. 17. For this calculation the magnitude of phase c voltage was assumed to be unity. The curves show that a considerable variation in torque with fault resistance may be expected. For hot stator and rotor windings the most unfavorable resistance appears to be at 0 pu. In an actual system, the exciter may not be capable of providing the excitation current needed to hold the V_c voltage up to 1.0. A calculation of the excitation under the conditions of fault resistance = 0 and hot stator shows that 5.2 units of excitation would be required. At an elevated temperature, this level of excitation may not be

available. For example, if a static-type exciter is designed to provide sufficient excitation to produce a 3-pu 3-phase short-circuit current when the rotor is at 250 C, then, at high rotor temperature and for fault resistance of 0 pu, the phase c voltage would be maintained at only 0.77 pu, with a 40% reduction in generator torque.

For the second example, consider the case of a line-to-line fault on a fully loaded generator. Fig. 18 shows the connection of the schematic sequence networks. The fault is assumed to be between phases b and c, with balanced static 3-phase rated load impedance on the system. From Fig. 18, the pu power may be written

$$P = R_{a1}I_{a1g}^2 + R_{a2}I_{a2g}^2 + \frac{V_{a1L}^2}{R_{1L}} + \frac{V_{a1L}^2}{R_{2L}} + I_f^2\frac{R_f}{3} + (R_{a2} - R_{a1})I_{a2g}^2$$

Again it is convenient to express all the voltages and currents of the power equation in terms of one of the phase voltages, say V_a.

From reference 2, the equations become

$$I_{a1g} = \bar{V}_a\left[Z_p\left(1 + \frac{Z_1}{R_f + Z_2}\right)\right]^{-1}$$

$$I_{a2g} = -\bar{V}_a Z_2\left\{(R_f + Z_2)Z_{2g}\left[1 + \frac{Z_1}{R_f + Z_2}\right]\right\}^{-1}$$

$$\bar{V}_{a1L} = \bar{V}_a\left[1 + \frac{Z_2}{R_f + Z_2}\right]^{-1}$$

$$\bar{V}_{a2L} = -\bar{V}_a Z_2\left\{(R_f + Z_2)\left(1 + \frac{Z_1}{R_f + Z_2}\right)\right\}^{-1}$$

$$I_f = \sqrt{3}\bar{V}_a[R_f + 2Z_1]^{-1}$$

These equations may now be evaluated for known values of the machine parameters and assumed values of R_f and V_a. The symbol

$$Z_p = \frac{Z_{1L}(R_f + Z_2)}{Z_{1L} + R_f + Z_2}$$

Z_2, as before, is the total negative sequence impedance viewed from the fault point. It is equal to the parallel combination of Z_{2g} and Z_{2L}.

The results of a calculation for a 20-kva generator are shown in Fig. 19. For this calculation it has been assumed that $V_a = 1.0$. If the excitation system is not capable of maintaining the phase a voltage at 1.0, then the torque is reduced by the square of the phase a voltage. For example, for a certain static exciter capable of supporting 3-pu 3-phase fault current at an elevated rotor temperature of approximately 250 C,

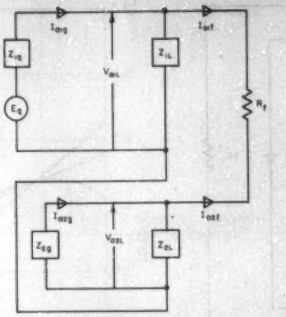


Fig. 18. Symmetrical component diagram of fault condition (line-to-line)

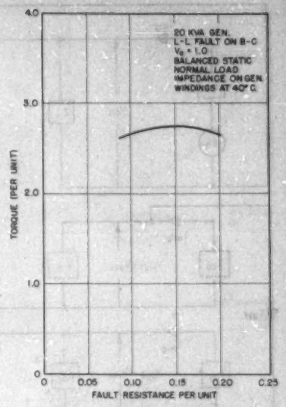


Fig. 19. Fault torques (line-to-line)

the curve of torque versus rotor temperature would be expected to appear about as shown in Fig. 8. The curve falls off above 50 C due to the inability of the exciter to supply sufficient excitation.

Discussion of Method

1. In order to get the actual torques imposed on the drive, to the calculated values must be added the losses described in parts 3 and 4 of the assumptions. These losses may amount to about 0.2 pu, so that the drive torques are somewhat higher than shown on the curves. A check on the foregoing calculations by measurement on an actual setup showed that the calculated internal generator torques were about 15% less than the measured values. This difference, in addition to expected errors in calculation and measurements, is caused principally by the harmonic losses together with the other losses which are not accounted for by calculation. It seems reasonable, therefore, in order to bring the calculations into closer agreement with actual values, to increase the calculated generator torques by 15% and this has been done in Fig. 8. It is clear that this 15% figure would be a function of generator terminal conditions, so that it is at best only approximate. However, in view of the present uncertainty of harmonic loss calculations, it would seem acceptable to rely upon this empirical factor. This value of 15% has been checked on two systems and gives good agreement on both.
2. Since under unbalanced conditions some dissymmetry in the machine will be called forth by local saturation and heating, an exact sequence theory would not permit the segregation of sequence networks as shown. However, since these effects are small and are very difficult to take into account, and in view of the uncertainty of the machine constants and harmonic losses, it is thought that the method applied here is adequate. However, if machine sequence constants are measured by use of standard fault tests where terminal quantities are near actual operating quantities, some degree of consideration is given to harmonic quantities since the constants are then determined under conditions of high harmonic generation.
3. If the faults investigated here should appear at the end of lines connected to the generator terminals, then the sequence impedances, viewed from the fault point, would be significantly increased, with consequent reductions in both torque and fault current.

References

1. Synchronous Machines: Theory and Performance (book), Charles Concordia. John Wiley & Sons, Inc., New York, N. Y., 1951.
2. Circuit Analysis of A-C Power Systems, Vol. 1, Symmetrical and Related Components (book), Edith Clarke. John Wiley & Sons, Inc., 1943.

Discussion

R. E. Smith (Westinghouse Electric Corporation, Lima, Ohio): The subject paper brings to light problems which perhaps have not been sufficiently investigated in the past, either by the electric system manufacturers or airframe manufacturers. It has probably not been recognized that class A generator systems can impose steady-state torques of 4 or 5 pu under conditions of short circuit at cold conditions. Knowing that high torque problems can exist with certain types of class C generators, perhaps engine pads and drives could be designed to permit the high torques without too much increase in weight. Any reduction in the generator torque must of necessity sacrifice the fault-clearing and motor-starting characteristics of the generator.

If the drives and engine pads cannot be strengthened to withstand the torques without too great a weight penalty, then perhaps the design of the generator-regulator combination should be reviewed before system performance is sacrificed. The authors make the general statement that class C system torques are sizably larger than is the case with class A systems and devote a large part of their paper to discussion of how these high torques come about and of methods of prevention. Admittedly, their statements apply to conventional a-c generators with integral d-c exciters and to static-excited a-c generators but the statements do not hold for brushless a-c generators with integral a-c exciters. It has been determined that brushless a-c generators

with integral a-c exciters which meet all class *C* requirements will, under any condition of short circuit, require no more torque than a class *A* system at −55 C. In fact this type of generator has a relatively small change of torque requirements from the hottest to coldest conditions. The reason for this is as follows.

The torque imposed by any generator is approximately proportional to the main generator field current. Thus the factors which limit this field current will also limit the generator torque.

First, let us take either a conventional a-c generator with an integral d-c exciter or a static-excited generator. Assume that for either type a short circuit exists at the hottest class *C* generator condition and that the ceiling excitation results in a steady-state torque imposed by the generator of 4.0 pu. Now let the generators be taken to −55 C. The generator field resistance will have been reduced by a factor of at least one half. For either of these types of generators the field current will be limited by the field voltage. Now, since the field resistance has been cut in half, the field current can be doubled to result in the same field voltage. Since the field current has been doubled, the torque imposed by the generator will have also approximately doubled or it is now 8.0 pu. Saturation has been neglected but it will also be neglected in the following comparison. If it were taken into account, it would show that even a greater range of torques is obtained from the foregoing types of generators than the one to be considered next.

Now consider a brushless a-c generator with an integral a-c exciter. An a-c exciter can be designed to have a high synchronous reactance, i.e., the excitation required by the exciter can be made by design to be much more dependent on its load current than its load voltage, the load current being the main generator field current and the load voltage being the generator field voltage. In fact, the a-c exciter can be assumed to be a current amplifier.

Now assume that the brushless generator is at its hottest class *C* condition and that a short circuit exists with ceiling excitation such that the steady-state torque imposed by the generator is 4.0 pu. Now take the generator to −55 C. The generator field resistance will have again been reduced by a factor of at least one half. If the generator field current remained constant, the field voltage will be cut in half. Naturally the generator field current must increase some with the same fixed ceiling excitation but it can only increase a small amount because, as mentioned, the excitation required by the a-c exciter is mainly dependent on the generator field current. Thus the fixed ceiling excitation can support only a small increase in generator field current. An increase of 20 to 30% would be typical. Thus the torque imposed by this type of generator at the coldest condition will be limited to 130% of the torque at the hottest condition or 5.2 pu compared to 8.0 pu of the previous types of generators discussed.

Therefore, the authors' conclusion that drive and engine pads cannot be designed to withstand maximum steady-state torques that a Class *C* alternator can develop without serious weight penalties, is not valid for all types of generators.

The paper has contributed to a better understanding of the different types of generator torques. I feel that the information given in the paper will help to place torque requirements in future specifications where they well should be.

S. C. Caldwell, T. E. Coppinger, and R. T. Smith: We appreciate Mr. Smith's interest in and comments on our paper. However we are not entirely in agreement with the discussion.

Reducing generator torques does not necessarily reduce fault currents beyond the minimums obtained under hot generator conditions. We have investigated this reduction and find that limiting total torque to 2 pu results in a surprisingly small reduction in fault-clearing capability (over only a limited range of fault resistance) in the 3-phase case and no reduction in unbalanced fault currents in the system studied.[1] Motor starting is usually characterized by low power factor. Hence in some of our tests with a large motor, we have found excitation ceiling limits starting current while drive torque is small compared with maximized fault torques. Thus, if torque limiting is used, motor starting characteristics are not likely to be affected. Obviously, if excitation is arbitrarily assigned a low ceiling in order to reduce fault torques, system overload and fault clearing performance will be adversely affected. This is not really torque limiting and is not the solution the authors have in mind. As a result of the foregoing and the fact that torque limiting appears feasible with but small penalties in weight and complexity, the authors feel that designing the drive, engine pad, and engine gears to accept torques of two or more times usual system overload requirements of 200% is unwarranted in most applications.

It was our intention to present a general method of analysis and not to compare various excitation systems. Obviously, these general methods can be used for comparison purposes, but a full torque study such as that presented in the paper in addition to analyses of transient response, complexity, reliability, and weight is required if valid comparisons are to be made. We think Mr. Smith's brief analysis is not comprehensive enough to form a basis for comparison and, further, it appears that significant factors are not considered in the particular case presented in the discussion.

Figs. 20–22 show the way excitation is obtained under 3-phase fault conditions in the two cases.

It is seen that the regulator is not a part of the loop in the static system since it only controls exciter performance without supplying any power entering into the excitation loop.

In both cases, the excitation systems are bootstrap systems and will build up to an excitation level determined by saturation somewhere within the loop. Thus, effectively, saturation is the limiting factor both in the static system and the brushless system.

In the brushless system, saturation in the current transformers supplying the regulator is the most probable limiting factor under 3-phase fault conditions. If this is the case, then the regulator magnetic amplifiers will be turned full on and the regulator approximates a constant voltage source over a sizeable range of exciter field current. Thus, we may replace the blocks with the following equivalent diagram.

V_R = the constant voltage output of the regulator (under fault conditions)
I_e = exciter field current
I_f = alternator field current
K = ratio between excitation current and internal voltage of the exciter
R_e = exciter field resistance
R_f = equivalent alternator field resistance
X_s = internal impedance of the exciter

Then, alternator excitation becomes:

$$I_f = \left| \frac{KV_R}{R_e(R_f + jX_s)} \right|$$

Now, in Mr. Smith's discussion it appears that the relationship between R_f and X_s is such that, over the field temperature range the quantity

$$\alpha = \left| \frac{KV_R}{R_f + jX_s} \right|$$

varies over a range of about 30%. Let us say then that

$$I_f = \frac{\alpha}{R_e}$$

The exciter field does not vary as greatly in temperature as does the alternator field since, usually, self-heating is less. Assuming that self-heating gives a rise of 30 C above cooling air temperature, we find in a class *C* system that exciter field temperature varies over a range of −55 C to +150 C and, therefore, exciter field resistance varies over a range of more than two to one. Thus, the total range of alternator

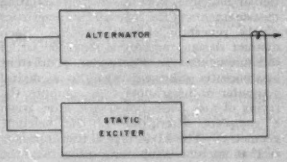


Fig. 20. Static exciter system

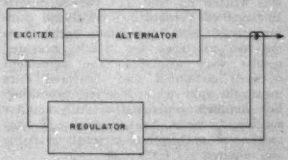


Fig. 21. Brushless (and d-c exciter) system

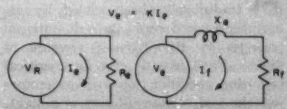


Fig. 22. Equivalent circuit of brushless exciter regulator

field current should approximate the product of the ranges of exciter field current with constant regulator output voltage and alternator field current with constant exciter field current. Then,

$$1<\frac{I_f}{I_{fb}}<2.72$$

where

I_{fb} = ceiling excitation with a hot machine

If we consider the static system, we find that the change in alternator field resistance is the only signficant factor as pointed out by Mr. Smith. In this case, over the range of −55 C to 250 C (including self-heating in a class *C* application) alternator field current varies as follows:

$$1<\frac{I_f}{I_{fb}}<2.76$$

Thus, the range of excitation over the temperature range is quite comparable in both cases. Using the simplified assumptions of the discussion we might expect an alternator which produced 8-pu torque in a static excited system would be benefited by use in a brushless system by a reduction in maximum possible fault torque to 7.9 pu.

As noted in Fig. 8 of our paper, the type of regulator sensing is a highly important factor when unbalanced faults are considered and, as shown, unbalanced fault torques with average sensing can be very much larger than with highest phase sensing.

As a result of these factors, we cannot concur with the analysis or the conclusions of the discussion. We think a much more comprehensive and detailed analysis is required before useful comparisons can be made.

Reference

1. Drive Load Requirements. *GER-1517*, General Electric Company, Schenectady, N. Y., 1958, pp. 37-39.

Prediction of the Thermal Behavior of Blast-Cooled Generators by Use of a Digital Computer

RICHARD M. MORONEY
ASSOCIATE MEMBER AIEE

Synopsis: The electrical rating of a blast-cooled aircraft generator depends on the flow rate and the temperature of the cooling air supplied. At the present time the nature of this dependence, for any particular machine, is obtained by correlating measurements of the thermal characteristics of the generator of that machine. Each distinct design requires a separate set of measurements. In this paper a different approach is discussed whereby a digital computer is programmed to compute the rating of any blast-cooled generator, given a complete description of the machine geometry. The advantage of this approach is that such a computer routine may be used to predict rating data for machines in the design stage.

A sample routine of the type proposed was written for the MIT (Massachusetts Institute of Technology) Whirlwind computer, and used to compute rating data on two machines for which measured results were available. The agreement between computed and measured data makes the approach appear very promising. For complete generality of the computer routine, however, heat-transfer equations more accurate than those presently available are required.

AIRCRAFT GENERATORS have traditionally been cooled by atmospheric air forced through internal passages by the dynamic pressure of flight. Until recently, the most critical operating condition for such generators was the taxi or stanstill part of a flight, when dynamic pressure was absent and operation was substantially self-ventilated. Increases in flight speeds to the supersonic range, however, have changed this situation in that now the highest-speed flight condition generally represents the most critical generator environment from a thermal viewpoint. This is because of the ram or stagnation effect, which produces a temperature rise, proportional to the square of flight velocity, in air taken aboard an aircraft. When this temperature rise becomes large enough, the improved heat-transfer coefficients afforded by the dynamic pressure of flight notwithstanding, generator temperatures exceed those associated with self-ventilated operation.

While most aircraft have not yet reached the speeds where a cooling system is necessary, many are flying at speeds where the specification of an appropriate generator requires a knowledge of the relationships which must be maintained between generator air flow, cooling air temperature, and electric load, in order to maintain reliable operation. The results of many careful investigations, aimed at determining the type and amount of measured data needed to specify the thermal characteristics of aircraft generators, were used in reference 1 to develop a rating method which is in general use at present.

The basic theory of reference 1 is data correlation: i.e., for certain specific inlet-air temperatures and pressures, necessary generator air flows are measured experimentally (by requiring some "critical hot spot" to be at a limiting temperature) and, simultaneously, data are taken on the pressure drop required to sustain this flow, the electrical operating point, etc. The data obtained are used, in conjunction with an assumption of appropriate forms for the governing air flow and heat-transfer equations, to obtain the constants in those forms which permit recalculation of the data points with minimum error in some sense, usually least-squares. The equations so obtained are then used to extrapolate to nonmeasured points by a graphical method. Figs. 1 and 2 are rating charts prepared by this correlation approach. Their use is described in reference 1.

The following expresses the idea behind the approach reported in this paper: Present theory in fluid flow and heat-transfer supplies adequate equations for calculating the thermal behavior of each component of a generator. By programming a computer to solve these separate problems, and "match the boundary conditions," i.e., find solutions for each part which are mutually compatible, one can obtain a computer model of a generator, as accurate as the state of the art permits, which may be con-

Paper 58-919, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted January 7, 1958; made available for printing April 16, 1958.

Richard M. Moroney is with Massachusetts Institute of Technology, Cambridge, Mass.

The author wishes to acknowledge the valuable assistance of A. Kusko, J. Duane, and R. Zehe, and also to express his appreciation of the assistance of the General Electric Company, Jack & Heintz, Inc., and the Eclipse-Pioneer Division of Bendix Aviation, in obtaining rating data and machine designs. The work described was performed under the sponsorship of the U.S. Air Force, Air Research and Development Command, WADC Contract no. *AF 33(616)-2921*.

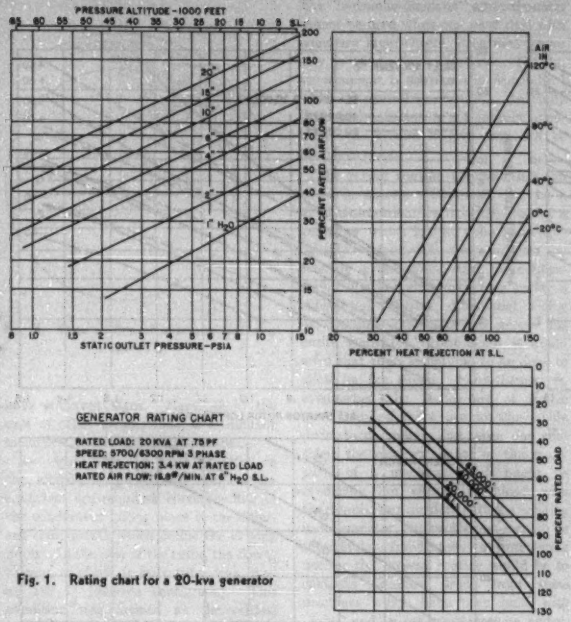


Fig. 1. Rating chart for a 20-kva generator

verted to a model of some different machine simply by changing a few geometric and electric constants. The accuracy of such a model is immediately verifiable by comparison of the results obtained experimentallly on any machine with those predicted by the computer. (Strictly speaking, the fact that such a model correctly predicts experimental rating data is not proof that the model is correct. Conceivably the model could follow an erroneous scheme of calculation and, by luck, arrive at a correct result. Practically, however, this possibility seems negligible, since it is known that the separate equations being solved are each at least approximately correct for the part of the generator they are simulating.)

The Sample Routine

Description

In accordance with the foregoing idea, a model was constructed which conformed to the following idealized description of a contemporary blast-cooled generator; see Fig. 3.

Air enters the generator from a duct (section *A*, Fig. 3) and is diffused by a blastcap, arriving inside the machine at section *B* with substantially zero velocity. Between sections *B* and *C* there is a heat input from the rotor end turns, which have a good heat-transfer coefficient because of their rotational speed. The air now passes through the alternator proper, a cross-section of which is shown in Fig. 4. Three distinct flow paths are available: one through the airgap, a second (actually a set of passages) between the rotor poles, and a third (also a set) behind the stator. The division of flow between these paths is assumed to be such that all exit pressures (section *D*) are the same. The air temperature at section *D* is a weighted average of the "path-exit" temperatures, if sufficient mixing is present. Heat is now absorbed from the downstream end turns and the cooling air exhausts through a large exit manifold.

The computer model which was prepared simulated each of the foregoing physical events by an appropriate equation. Evidently there are two separate sources of error in such an approach. First, it will have been noted in the preceding paragraph that only certain effects, presumed dominant, are considered. This idealization of a generator as a "lumped parameter device" can cause errors arising from the neglect of certain cooling mechanisms and air-flow phenomena, and additional errors due to forcing inexact boundary conditions on the equations used to simulate the "dominant" effects. The second source of error is in the parameter equations themselves, which are only approximations to the equations truly descriptive of the complicated physical events taking place.

Reference 2 gives a careful discussion of the errors just mentioned, as well as a more thorough treatment of all the material in this paper. Only the most significant errors are described in the following paragraphs.

Inaccuracies in the Sample Routine

The routine described is basically a computation of the air flow distribution in the alternator section of a generator. The assumption is that if cooling-air flow can be properly computed for each duct in an alternator, then it is a simple matter to put a rating on the machine in question. In the case of so-called tight generators, i.e., generators which operate with low air flows and high pressure drops, the preponderance of air flow events will indeed take place in the alternator section. Otherwise the design would be a poor one, since presumably the pressure drops in other sections can be held to the same level as those in a loose machine, and these will be a smaller fraction of the total drop in the tight machine. (Note that the presence of pressure drop is a direct indication of the presence of "air flow events," in particular, heat transfer.)

In loose machines it is not reasonable to expect alternator section phenomena to predominate, since "stray" pressure drops become sizable fractions of the total drop in such machines. The approximation should become more reliable, however, at operating conditions of high air flow and low density, since pressure drop increases sharply under such conditions, whereas most stray effects are more or less constant; some may become smaller with decreasing density, for instance, fan effect due to end turns.

Once the routine has computed the air-flow distribution in the alternator ducts, it proceeds to calculate the heat removed from the rotor, under the assumption that this equals the heat absorbed by the cooling air in the interpole ducts and the air gap. This assumption implies that the heat transferred from the stator into the air gap is exactly equal to the heat transferred from rotor to stator, which may be considered a compromise, since in different machines the radial derivative of stator temperature has been found to

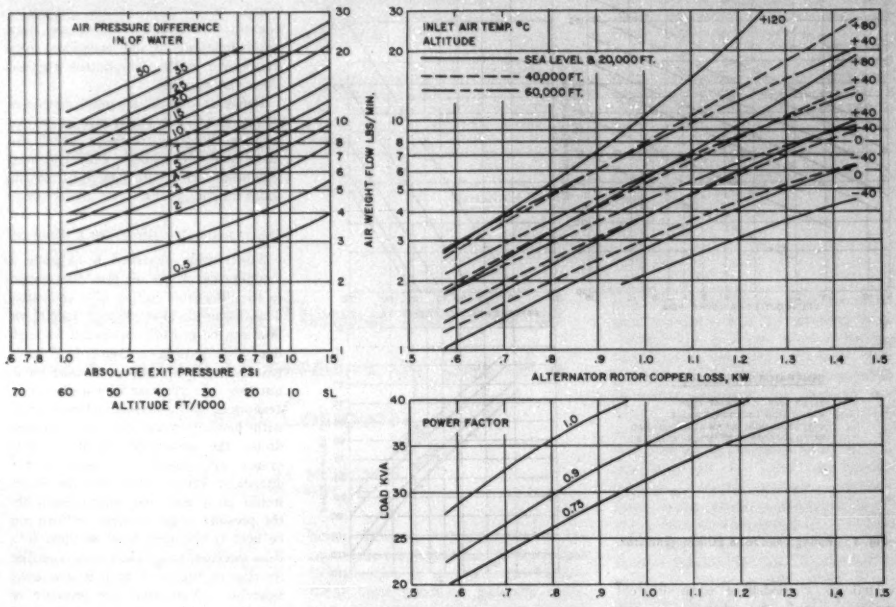


Fig. 2. Rating chart for a 30-kva generator

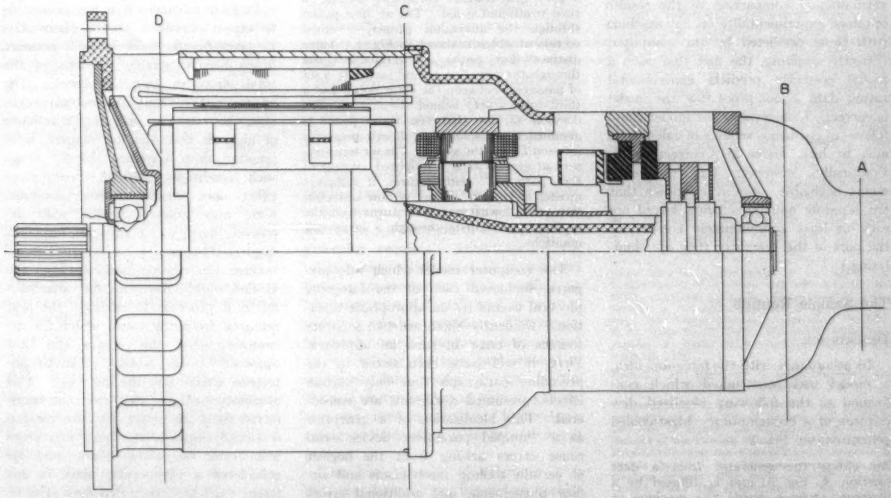


Fig. 3. Schematic of air flow paths in a generator

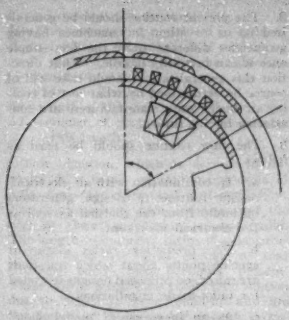


Fig. 4. Cross section of alternator part of a generator

have different signs. Here, as in the case of other assumptions, it is difficult to suggest a more rational approach.

In computing the heat transfer in the alternator ducts, wall temperature equations are required, corresponding to the conduction taking place in the different components which define the cooling ducts. In the case of the rotor, the downstream winding temperature was fixed at 285 C (degrees centigrade). This condition was defined as the critical thermal situation, in accordance with the fact that the rotor winding is the critical thermal component for contemporary machines under most conditions. It was deemed within the expected accuracy of the computation to fix the wall temperatures of the air gap and the stator ducts at representative values.

In computing the air flow through cooling ducts, the actual flow condition was approximated by equations valid for fully developed turbulent flow. This approximation is discussed in Appendix I. Briefly, it amounts to neglecting the boundary layer effects at inlet to a duct and considering conditions to be everywhere the same as they are eventually, far downstream, when the flow is in a steady state as seen by an observer moving with flow velocity, i.e., when radial fluid motions have disappeared. Computed results have indicate that this approximation may not be valid, since the entrance region of some ducts extended for half the duct length. It is impossible to improve this approximation with the present state of the art of fluid flow computation, however; theory is well developed for the case of round tubes, but there is no literature on the case of irregularly shaped ducts such as those found in aircraft generators. The "hydraulic diameter" approximation cannot be used when one must deal with boundary layer effects, since then geometry is of crucial importance. Finally, an argument is advanced in Appendix I which shows that, whereas the fully developed equation is not entirely applicable, the results it give may not be greatly in error.

The rating routine assumes that fan effect is absent. This may not be a valid assumption, even in the case of a generator which has no component specifically introduced for purposes of creating a fan effect, since the configuration of rotor end turns may create a substantial pressure differential aiding or opposing that provided across the entire generator. Here again, it would seem that there is nothing to be done to allow for fan effects. Fans themselves, even when the blades are of simple aerodynamical shape, are not thoroughly understood devices, and when one considers the random shapes of the many parts of a machine which might be producing fan effects, combined with the complicated air-flow patterns existing in a generator, the problem is seen to be formidable. Incidentally, an interesting use for the present routine would be to compute pressure drops from some machines which have been built with "fans" and to see whether these embellishments were justifying their use.

The accuracy of the assumption used to determine the air flow distribution between different alternator passages, i.e., the assumption that the pressures at the exits of each passage are the same, is seen to depend in large part on the previous paragraph. If a fan is present it is to be expected that there will be a radial difference in pressure of some magnitude, along with the axial differential. However, in order to perform the computation some constraint is necessary, and the present assumption appears to be the only reasonable one.

Two final inaccuracies are related to the lengths used for different cooling-air passages. First, air-flow effects due to the rotation of the rotor were crudely approximated by lengthening the interpole and air-gap ducts by a factor which was the ratio of vector velocity magnitude to axial velocity magnitude, the vector velocity being the vector sum of the axial and rotational velocities in a particular duct. Reference 3 provides some experimental justification for this approach. Finally, it was found after completion of the routine that stator duct length should be different from the others in order to allow for the fact that air intended for the stator ducts must flow around the stator endturns, which are so interleaved as to form a continuous barrier. This correction could not be made directly without rewriting the entire program, so the stator duct cross-section area was decreased slightly instead. In any case, this did not affect the computed results.

The intention of this lengthy section on inaccuracies is not to degrade the routine prepared. The preceding assumptions and approximations are listed in order to give the reader a realistic estimate of the reliability of results computed by the routine, but it is to be emphasized that all of the assumptions and approximations used, however gross some may seem, are the best that can be found with the art in its present state. Until such time as extensive experimental work has been completed, few significant improvements in this routine will be possible. It is hoped, however, that the existence of this routine will serve as a stimulus to the needed experimentation.

Results Computed by Sample Routine

Table I gives the results obtained by using the computer routine of this report to predict rating data for the generators whose rating charts are Figs. 1 and 2. The first three items on any line of the table (after the identifying letter) represent the cooling-air conditions fed into the computer. Additional input consisted only of geometric and electric constants, as listed in Appendix II. The last three items of any line of the table are the results supplied by the routine, namely, the pressure drop required to obtain the specified air flow, the expected exhaust-air temperature, and the permissible rotor loss for this cooling condition, the latter expressed in per unit of rated rotor loss. Pressure drops and rotor losses obtained from the rating charts are given in parentheses.

Lines *A* and *B* for Fig. 1 and for Fig. 2 in Table I show the present routine at its best. The coolant conditions of high air flow and low inlet pressure cause those effects considered by the routine to predominate, and hence give the best chance for accurate results. Nonetheless, when the assumptions and approximations of the previous section are considered, the results are truly gratifying. For both generators the losses are underestimated by 0.1 per unit, which indicates, as expected, that certain neglected cooling mechanisms are not completely absent. For the generator of Fig. 2 the pressure drops are computed

Table I. Computed Ratings for Generator of Figs. 1 and 2

Line Identification	Air Flow, Pounds per Minute	Inlet Pressure, Pounds per Square Inch Absolute	Inlet Temperature, C	Pressure Drop Inches, H_2O	Exhaust Temperature, C	Per Unit Rotor Loss
			Fig. 1			
A	16.8	6.0	120	12 (15)	144	1.3 (1.3)
B	8.4	2.5	100	8 (9)	131	0.8 (0.9)
C	16.8	14.7	120	3 (6)	144	0.9 (1.3)
			Fig. 2			
A	10	3.5	60	13 (13)	117	1.1 (1.2)
B	10	3.5	120	13 (13)	152	0.8 (0.9)
C	10	14.7	120	2 (3.5)	152	0.3 (1.0)
D	5	4	100	3 (3)	148	0.4 (0.8)

as accurately as the rating chart can be read, while those for the generator of Fig. 1 are underestimated slightly. Again, the error in the case of the latter machine was to be expected, since that generator is of loose design (offering relatively little restriction to air flow) and therefore the alternator section effects can never completely dominate the pressure drop picture.

Line *C* for Fig. 1 and Fig. 2 in Table I shows the reduction in accuracy which results when the ratio of air flow to inlet pressure is reduced. Actually, it appears that the inaccuracy under these conditions is mostly due to the use of fixed-point representations for Mach numbers and velocities in the air-flow routine. Very small changes in an initially small Mach number are not resolved by the routine, and small inlet velocities are not computed to a sufficient number of significant figures to allow for accurate computation of inlet static pressures by equation 8. This difficulty can be cured immediately by using floating-point numbers for all air-flow computations. In any case the original motivation for using fixed-point numbers proved invalid.[2]

The accuracy of both lines *A* and *B* of both lines in Table I can be improved simply by embroidering the present routine with small additions designed to take account of secondary effects (such as conduction down the shaft to the bearings and thence to the cooling air).

Lines *C* and *D* for Fig. 2 in Table I illustrate a final difficulty. The generator of Fig. 2 does not have separate interpole and air-gap passages, as suggested by Fig. 4. Instead, the separating wedge shown in that figure is absent. Recent careful studies have shown that for such a configuration the rotor heat-transfer coefficients will be very much larger than predicted by the present routine.[4] The difficulty here, then, arises from attempting to fit the computer routine to a generator for which it is not general enough to approximate. The correction, again, is to enlarge the present routine, after development of the necessary theory. It is interesting to note that the effect referred to does not appear in lines *A* and *B* for Fig. 2 of Table I, which indicates that it is essentially a high-density phenomenon.

To summarize the results of testing the sample rating routine:

1. For the conditions it could be expected to handle the routine was excellent.

2. Changing the air-flow computations to floating point would extend the conditions of item 1 to the entire range of interest in most cases.

3. The routine is not general enough in its present form to cover all existing generator configurations.

Conclusions

1. Digital computation of the steady-state air flow and temperature behavior of blast-cooled generators is feasible, and with some further experimental work such computation should be extremely accurate.

2. Use of a digital computer for obtaining rating data on *existing* machines may not be economically desirable, so it would seem that the importance of computers will be in predicting rating data for proposed machines. [The present routine required about 5 minutes of computer time (the cost being $20 on the MIT Whirlwind) to compute rating data for one cooling-air condition. Measurement of a set of 12 points is said to cost about $80, which makes the measured data appear cheaper. There is no way, however, to measure temperatures on a blueprint.]

3. Medium-speed computers such as the IBM *650* may suffice for the type of work done for this report on the high-speed MIT Whirlwind, since certain considerations in Appendix I indicate that a shorter routine will be just as effective as the present one.

Recommendations

1. Experiments should be undertaken which will permit removing the inaccuracies and omissions of the present routine in so far as possible.

2. The present routine should be generalized so as to allow for machines having geometries different from the few simple ones it can presently handle. What direction this generalization should take will of course be determined by what sort of cooling-air ducts, etc., manufacturers are considering for the future.

3. The new routine should be used as follows:

a. In combination with an electrical-design routine to design generators optimum from the thermal as well as the electrical viewpoint.

b. For predicting rating data, at specific points about which questions are raised, on proposed designs intended for tailor-made installations.

c. As an inexpensive "model shop" which permits designers try out their ideas about cooling passages without building and testing a real machines.

d. As one component in a larger computer routine capable of studying the behavior of a combined generator and cooling system.

Appendix I. The Air-flow Computation

At the start of the present research it was anticipated that many of the nonlinearities found on rating charts could be explained by the fact that the correlation equations used in constructing the charts are based on incompressible flow. For this reason the routine used a rather high-powered equation for computing pressure drops and heat transfer, one which takes into account the effects of compressibility. In retrospect this appears to have been a waste of effort. It is true that, when iterating to arrive at a proper distribution of air flow among the ducts, the computer frequently assigned to a duct an air flow of such a nature that compressibility effects were indeed present. The final air-flow distribution, however, was always such that incompressible flow equations would have provided satisfactory accuracy. If this routine were to be reprogrammed it might well pay dividends to delete the compressible flow allowances and, instead, concentrate on including a routine which could allow for the entrance region effects which will be mentioned presently.

The air-flow computation is a step-by-step numerical integration of the differential equation

$$\frac{\Delta M^2}{M^2} = F_T(M^2)\frac{\Delta T_0}{T_0} + F_f(M^2)4f\frac{\Delta x}{D} \qquad (1)$$

where

$$F_T(M^2) = \frac{M^2(1+KM^2)\left(1+\frac{K-1}{2}M^2\right)}{1-M^2}$$

$$F_f(M^2) = \frac{KM_4\left(1+\frac{K-1}{2}M^2\right)}{1-M^2}$$

and

K = air exponent, a constant
M = Mach number at position x in a duct
T_0 = stagnation temperature at the same position, degrees Fahrenheit absolute
f = friction factor
D = hydraulic diameter of the duct
Δx = length of the step being taken
ΔT_0 = change of stagnation temperature in this length

This equation, which is derived in reference 5, is based on the assumptions of fully developed flow and turbulent flow. The computations made indicate that the first of these assumptions is not valid. The length-to-diameter ratios of generator-cooling ducts are small enough so that entrance effects are probably present. These effects result from radial fluid motion due to boundary layer build-up and are not found at locations far enough downstream to have a "fully developed" boundary layer. In the case of the present problem, calculations based on the usual idealized conditions of smooth ducts and bell-mouthed inlets indicate that about one half of the lengths of the stator and interpole ducts are in the entrance region. Here an improvement in the routine would seem to be called for, but there is no fluid flow theory available at the moment which can cover the problems of interest (in particular, the problem of noncircular duct cross sections). Extensive research would probably be necessary to improve this phase of the routine. A compensating effect is that air at inlets to generator cooling ducts is certainly not "still" air and probably has a large amount of turbulence, which would accelerate the boundary-layer forming process and hence reduce entrance effects.

Entrance effects improve the local heat-transfer coefficient and increase the pressure drop, effects which could be approximated by an increase in the friction factor, but since the present routine already uses a large friction factor (as will be seen) it can be hoped that substantial errors are absent. It is to be noted that the diameter Reynolds Numbers in most ducts were just barely in the turbulent flow regime (approximately 3,000).

Equation 1 is not sufficient in itself for computing the pressure drop through a duct. An additional equation is required connecting the stagnation temperature of air at any section with the location of that section along the duct. A combination of the Reynolds analogy and the approximation that heat flux is constant along the length of a duct gives the following additional equations, also from reference 5:

$$T_w - T_0 = T_{w1} - T_{01} \tag{2}$$

$$\Delta T_0 = (T_w - T_0) 2f \frac{\Delta x}{D} \tag{3}$$

where subscript 1 refers to the duct inlet and T_w is the duct wall temperature at length x from the upstream end.

The Reynolds analogy is a relationship between heat-transfer coefficient and friction factor which is found to be quite accurate for highly turbulent flows. The friction factor itself is a function of the Reynolds number, so that in general it would have to be calculated at each step of the integration. Fortunately, the friction factor in rough pipes is almost a constant above a certain Reynolds number,[6] and thus may be taken as such in the ducts under consideration, i.e., passages defined by laminations and varnished insulation. To summarize the foregoing, in order to calculate the pressure drop in a duct it is necessary to have a friction factor, an expression for heat-transfer effects, and a knowledge of the type of fluid motion to be expected. A good approximation has been attained for the first of these factors, a reasonable approximation for the second, and for the third a tenuous approximation, because of gaps in present theories.

It will be noticed that equation 1 is in terms of the flow Mach number. This equation, in combination with equations 2 and 3, gives as output the Mach number and stagnation temperature in the exit plane of a duct. The same numbers will also produce the exit pressure by use of the following equation (which is one of those from which equation 1 was derived):

$$\frac{P_2}{P_1} = \frac{M_1}{M_2} \sqrt{\frac{T_{02}}{T_{01}}} \sqrt{\frac{1 + \frac{K-1}{2} M_1^2}{1 + \frac{K-1}{2} M_2^2}} \tag{4}$$

Equation 1 is in terms of the Mach number because that variable gives the most convenient form to compressible-flow equations.

The numerical integration of equation 1 was performed by a simple predictor-corrector formula in which the change in M^2 in a length Δx was first assumed to be the same as in the previous Δx, in order to compute an average Mach number for the section from x to $x+\Delta x$. The functions of Mach number on the right side of equation 1 were then computed, with the use of this predicted average Mach number, and equation 1 then supplied a corrected change in M^2, which was used to compute the Mach number at $x+\Delta x$. This cycle was iterated once. The step taken is given by

$$\Delta x = \frac{L/D}{[L/D]} D \tag{5}$$

where [ ] means "the integral part of." This step is seen to be close to one hydraulic diameter, and the number of steps required is precisely $[L/D]$.

Aside from the entrance effects already mentioned, which were due to the build-up of a boundary layer, there is an effect immediately at the entrance or exit of a tube resulting from the abrupt change in area.

If station 0 is taken as a point just outside the inlet of a tube, and station 1 a point just inside the inlet, the following equations hold:

$$V_1 = \sqrt{(T_0 - T_1) 2g J c_p} \tag{6}$$

$$\frac{W}{A V_1} = \rho_1 \tag{7}$$

$$\rho_1 R T_1 = P_1 \tag{8}$$

$$P_0 = P_1 + (K+1) \left(\frac{\frac{\rho_1}{2} + \frac{\rho_3}{2}}{2} \right) V_1^2 \text{ (empirical)} \tag{9}$$

where

Table II. Parameters on Data Tape

Symbol	Parameter	Units
Q_e	exciter loss	watts
W	air mass flow	lb/min*
T	inlet air temperature	C
P	inlet air pressure	psia†
r	mean radius to rotor end turn	inches
ω	speed of rotation	rpm
h	height of rotor winding	inches
T_R	estimated upstream rotor winding temp‡	C
ρ_R	number of poles	
A_R	cross section of rotor interpole duct	square inches
A^*	cross section of rotor winding	square inches
Z_R	one per-unit rotor loss	watts
δ	air gap	inches
L	cooling-passage length	inches
S_R	perimeter of interpole duct	inches
$2g$	mean air gap diameter	inches
T_G	estimated upstream air-gap wall temp	C
ρ_s	number of stator-cooling passages	
A_s	area of each stator passage	square inches
S_s	perimeter of each stator passage	inches
T_s	estimated upstream stator wall temp	C
A_g	area of inlet air duct	square inches

* Pounds per minute.
† Pounds per square inch absolute.
‡ Temperature.

ρ = density
g, J, c_p, K = constants

These equations may be combined to give the following formula for the velocity at station 1:

$$0.218X^2 + X^2\left(0.120 t_0 \frac{W}{PA}\right) - X t_0 + \left(0.89 t_0^2 \frac{W}{PA}\right) = 0 \tag{10}$$

where

$$t_0 = \frac{T_0}{1{,}000}$$

$$X = \frac{V_1}{1{,}000}$$

This polynomial has two real positive zeros, because equations 6 through 9 take no account of the second law, which is needed to specify the inlet velocity completely. The second law says here that the Mach number can not exceed unity at station 1. An original oversight of this trivial fact caused considerable delay in program debugging.

Appendix II. Inputs for Sample Routine

The choice of independent variables was such that the rating routine provided a generator power rating and a required pressure drop, given the mass flow and inlet state of the cooling air. The parameters on the data tape were as shown in Table II.

The significance of most of Table II is obvious. A word about two parameters, however: r, the mean radius to the rotor end turn, is a number used to find an

average velocity of the rotor end turns for purposes of calculating a heat-transfer coefficient from end turns to cooling air; A^*, the rotor winding cross-sectional area, is not used by the sample routine. It was included in the expectation that eventually it would prove desirable to make the routine capable of providing a power rating in terms of rotor current density, as well as in terms of rotor loss (the former factor would seem to be more important to the designer).

References

1. Correlation of Rating Data for Rotating Electric Machinery, Erwin O. A. Naumann. *AIEE Transactions*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 443–52.

2. Digital Computer Rating of Blast-Cooled Generators, R. M. Moroney. *WADC Technical Report no. 57-522*, Wright Air Development Center, Dayton, Ohio (to be published).

3. Air Gap Heat Transfer in Rotating Electric Machines, Carl Gazley, Jr. *Laboratory Report DF52GL156*, General Electric Company, Schenectady, N. Y., July 16, 1952.

4. Surface Heat Transfer Coefficients of Salient Poles in a Blast-Cooled Alternator, F. S. Tse, W. Robinson. *WADC Technical Note 56-123*, Air Research and Development Command, Dayton, Ohio, Apr. 1956.

5. The Dynamics and Thermodynamics of Compressible Fluid Flow, Vol. 1 (book), Ascher H. Shapiro. The Ronald Press Company, New York, N. Y., 1953, chap. 8.

6. Measurement of Friction in a Pipe for Subsonic and Supersonic Flow of Air, J. H. Keenan, E. P. Neumann. *Journal of Applied Mechanics*, New York, N. Y., vol. 13, no. 2, 1946, p. A-91. (As quoted in reference 4.)

Computer Analysis of A-C Aircraft Generators

J. R. M. ALGER
ASSOCIATE MEMBER AIEE

E. F. MAGNUSSON
MEMBER AIEE

J. T. DUANE
ASSOCIATE MEMBER AIEE

R. T. SMITH
ASSOCIATE MEMBER AIEE

ACCURATE electrical performance data obtained quickly and early in the preliminary design stages are necessary for aircraft generator design. The nature of the business is such that relatively few machines are designed in a given time, but each must meet a quantity of difficult specification requirements. These engineering requirements provide a natural application of the stored-program digital computer. A performance analysis program was needed which would accurately, and quickly, provide complete data on conventional a-c aircraft generators, as dictated by system requirements. Such a basic program was defined in October 1956, work started in November, flow diagram completed in January, and the program was debugged and running in March 1957. Since that time all generator designs have had a program run. These runs provide complete and accurate information for the designers and users of the machines. All new designs obtain the benefits of such early and detailed data.

This paper discusses the unusual aspects of program use and flexibility, considering the output data desired and the special forms of data presentation required. Some of the machine theory involved is discussed, as is the use of computer logic circuitry. Also, the general results obtained from using the computer in generator design are discussed, including the means by which the engineering organization handles the new problems created by the availability of this new tool.

Paper 58-908, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N.Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing April 30, 1958.

J. R. M. Alger and E. F. Magnusson are with the General Electric Company, Schenectady, N. Y.; J. T. Duane is with the General Electric Company, Erie, Pa.; and R. T. Smith is with the University of Texas, Austin, Tex.

Program Output Data

At the start of the project, much thought was given to the determination of what output information would be desired. All the possible modes of operation in an aircraft were envisioned and the program was designed to calculate generator performance under each of these conditions. The list of output data developed is given in Table I. These data are required not only by the generator designer, but also by the engineer responsible for design of the control and excitation components in order that a particular system can be developed in a minimum of time. For these reasons the controls engineer must be given generator data at the earliest possible moment so that the generator and controls can be developed together. Also, the systems engineer must be supplied with generator performance data so that under all possible modes of system operation he will be able to predict system performance. Therefore, the data obtained as outputs of this program are designed to supply the differing needs of the generator, controls, and systems engineers.

The results (see Table I, section I) are in two sections. The first section gives generator resistances, reactances, time constants, weights of copper and iron, mean length of turn, sustained fault currents, voltage unbalance, and other design parameters. These quantities are necessary for predicting various modes of operation. For example, unbalanced operation due to faults requires calculation of the fault currents by means of symmetrical components. Therefore, the symmetrical component impedances are required. A further example is that of voltage recovery after a shift in load. Transient reactance (X_D') determines the initial jump in voltage and the rotor open-circuit time constant (T_{DO}') is partially responsible for the time delay in obtaining steady voltage again.

The second section of print-out presents excitation calculation results over a wide range of possible operating conditions; see Table I, section II. Both normal and abnormal loadings are presented in sufficient detail to permit plotting curves relating excitation field current; terminal voltages; phase currents; rotor watts; flux densities in the rotor, teeth, and core; ampere-turns in the rotor, teeth, and core; and power factor. In this way the variety of electrical loading conditions on the generators encountered in system operation on the airplane are completely covered.

Table I. Program Output Data

SECTION I. PARAMETERS

1. Reactances: $X_A - X_D - X_Q - X_D' - X_D'' - X_D'' - X_Q'' - X_0 - X_2$
2. Resistances: $R_A - R_F - R_3 - R_{KD} - R_{KQ}$
3. Time constants: $T_A - T_{DO}' - T_{DO}'' - T_{QO}'' - T_D' - T_D'' - T_Q''$
4. Phase voltages for specified load unbalance
5. Fault currents for unsymmetrical faults

SECTION II. PERFORMANCE

1. Balanced loads, i.e., no load, full load, etc.
   Flux densities in various parts of the magnetic circuit, field current, and power.
2. Faults, i.e., 3-phase, line-to-line, single line-to-ground, etc.
   Flux densities in various parts of the magnetic circuit.
   Field current and power.
   Unfaulted phase voltages (unbalanced faults).

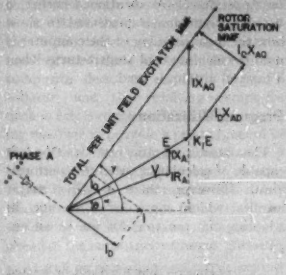


Fig. 1. Salient-pole synchronous-machine vector diagram

Program Input Data

The information needed to calculate the desired outputs may be divided into four basic categories of input information as follows:

1. Basic machine dimensions and characteristic curves for the magnetic material used.
2. Operating conditions under which generator performance is to be calculated.
3. Temperature at which the windings will be operating.
4. Range and type of output information required.

The 56 pieces of design data in category 1 are readily available from the design sheets and machine drawings. Effects of changes in any of these dimensions can be studied by changing the appropriate input value. Also, punched cards giving characteristic curves for various steels are necessary. One set of cards is selected on the basis of the desired steel specification.

The input data required in categories 2, 3, and 4 are affected by the application for the output data. The operating conditions, such as load, power factor, and voltage, are usually taken from a specification.[1] Likewise, the cooling conditions are usually specified and these determine the insulation system that must be used and the corresponding winding temperatures. Also, the generator application determines to some extent the range and type of output information required. In order to handle this situation, several decks have been established to handle problems of varying complexity and detail. This manner of obtaining flexibility of operation is discussed further under "Program Utilization."

Machine Theory

The theory of salient-pole synchronous machines has been steadily developed and refined over many years. However, in a large area of design the experience and judgment of the individual designers must be relied upon very heavily. As a result of individual judgment, there are many formulas available for calculating various machine characteristics, i.e., the leakage reactance of the generator, and the equations used in computing the performance of the a-c aircraft generator have been developed by many contributors. Nonlinear effects due to saturation and complex geometric configurations of magnetic circuits present formidable obstacles to precise mathematical solutions. However, the digital computer allows calculations too tedious for hand calculation to be done swiftly and precisely. Use of the computer allows saturation effects to be accurately brought into account.

The "constants" of the generator are calculated from the physical dimensions of the magnetic circuits and winding arrangements, using modified revisions of formulas available in the literature.[2–4] These formulas are used in calculating the output of section I in Table I, i.e., time constants, resistance, and reactance. Also included in this section are the fault currents (for unsymmetrical faults having rated voltage on the unfaulted phase) and the phase voltages with one third of rated load on one phase. These quantities are obtained by the methods of symmetrical components, using the calculated generator sequence impedances.[5–7]

The excitation calculations for section II output data are carried out by computing flux densities in the various portions of the magnetic circuit, entering

Table II. Floating-Point Number

Fixed Point Data	Floating-Point Representation
00.100	.491,000
01.000	.501,000
10.000	.511,000

magnetization curves for the corresponding field intensities, and summing the component ampere-turns over a path linking the field winding. As an example of these procedures, field excitation for balanced loads is calculated using the equations of Appendix I as derived from the vector diagram of Fig. 1. The treatment is not unusual except possibly in the way saturation effects are handled. The mmf (magnetomotive force) required to overcome saturation is divided into two components: one which is dependent only on the total flux, the other a function of the field mmf as well as the direct-axis flux. The first is represented by an mmf due to stator core and tooth saturation and the second by saturation in the rotor pole piece. This assumption is made possible by the fact that there is very little saturation in the quadrature-axis magnetic path in the rotor, i.e., in the pole tips. Most of the saturation in the rotor exists in the body of the pole, at and near the base, which is in the path of the direct-axis flux.

Under steady-state conditions, therefore, the rotor pole component of mmf is calculated as a function of the direct-axis flux and pole-face mmf. The value of stator core and tooth saturation mmf is represented by K_1 of Fig. 1 and is

DECK 302/3 PROBLEM 257 PAGE 01

J. DUANE 12/4/57 2CM&3A1 22KVA 4 POLES

STA. STL. B9E14D ROT. STL. B3C3A

OUTPUT - RESISTANCES AT 25 DEG. C

556 0005580007	X/AD 5017389125	X/AQ 4980311636	X/LF 4847770860	X/L SKEW 4759576821	X/L 4982733733	X/F 4919939621	X/KQ 4912208963
563 0005630007	X/KD 4893541018	X/D 5018216462	X/Q 4988595009	X/D1 4926161780	X/D11 4914415812	X/Q11 4917330964	R/KD 4880814922
570 0005700007	R/KQ 4896681072	R/F 4758308377	R/A 4818074252	R/E 4845768133	X/E 4915093813	X/KKD 5018324535	X/KKQ 4905520599
577 0005770007	X/FFD 5019363087	T/A 4735428990	T/D01 4913226720	T/D1 4918995705	T/DO11 4713412684	T/D11 4673906305	T/Q011 4737138177

VOLTS UNBALANCE, 1/3 LOAD, 3/4 P.F.

584 0005840007	T/Q11 4872657938	R/0 4830726228	X/0 4815887293	V/PH A 4998066548	V/PH B 5010001584	V/PH C 5010191470	THETA B 5224137156

VOLTS UNBALANCE, 1/3 LOAD, 1.0 P.F.

591 0005910007	THETA C 5212018617	V/PH A 4999150913	V/PH B 4999028469	V/PH C 5010182215	THETA B 5224176464	THETA C 5212129640	I/LL 5051760294

598 0005980007	I/B LLG 5120065615	I/C LLG 5116038574	W/3Q IN S 5019744438	W/3Q IN R	LB/TOTAL 5126774868	MLT/STA 5012531723	MLT/ROT 5010302380
605 0006050007	LB ROT FE 5110543720	LB/ROT CU 5027311891	LB/STA CU 5033231447	LB/STA FE 5099642947	LB/BARS 4921258637	I/SLG 5086110429	4914959965
612 0006120007	GMS PU 5019607843	P/T 4951589124	T/AE 5123184000	K/TURN 4712899949	P/BL 4956204782	K/SKEW 4999906778	P/BR 4934100048

Fig. 2. Partial presentation of output data

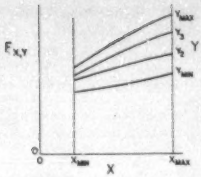


Fig. 3. Graphical function of two variables

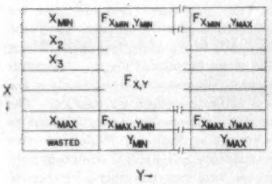


Fig. 4. Storage of function of two variables on computer drum

determined from a knowledge of the stator iron characteristics and the required induced voltage. When a tapered air gap is used, the effect of saturation is to "spread out" the flux across the pole face more and more as the saturation increases. The manner of handling this situation is discussed as a problem in the use of logic circuitry in Appendix II, since iterative techniques are involved.

The excitation requirements for unbalanced-load conditions are obtained by the use of symmetrical components applied to a method similar to the balanced-load procedure.

Programming

Programming was done for a specific electronic digital computer, namely the IBM (International Business Machines Corporation) *650*. While some of the available *650* computers had built-in floating-point and indexing accumulators, others did not. There were various *650* computers available with a variety of accessory equipment; and additions were pending to most of them at uncertain dates in the future. But, none available to the generator designers had core memory, instantaneous access storage, magnetic tapes, or disks.

Therefore, the programming was done with an older floating-point calculation procedure which, while slower, would permit running on the largest number of *650* computers and suit operating requirements at the earliest moment. Floating point is a descriptive term applied to data printed with the decimal-point position indicated by a number attached to one end of the data. Table II explains floating-point numbers, and Fig. 2 shows how the output data appear using this floating point.

One requirement of the program is the presence of 13 graphical functions, each of two variables. Also, a single function in graphical form of one variable is present. The graphical functions are the required flux coefficients and magnetic saturation curve data.

In this case, the capacity of the computer would have been overtaxed if conventional programming methods had been used. The program would have had to fit into leftover spaces after the tables of functions had taken the choicest storage elements. Because of this, a new procedure was designed that handles general functions of two variables. Fig. 3 shows a general funtion of this type and Fig. 4 shows how the data representing the function are stored on the drum of the computer. The program procedure obtains the value of the function by linear interpolation for any arbitrary value of x and y. Also, the procedure checks for errors. Each set of x and y values is checked against the table to make sure that it does not fall outside the table limits. If either x or y does so, the program stops, and at the same time indicating lights appear on the console of the computer.

A large number of logic operations (if not "this" then "that") were necessary in order to tell the computer how to take up in turn all of the specified load and fault cases. Fig. 5 shows a much simplified flow diagram of the entire program. Section I data (Table I) are obtained first. In order that the excitation calculation procedure may be followed, the machine terminal conditions, i.e., voltage, current, power factor, and balance, must be specified. This is done by a set of instruction cards which tells the program what "cases" (each set of machine terminal conditions constitutes a case) to calculate. It is this flexibility that allows derivative decks to be made up. Each derivative deck has its own particular set of instruction cards telling the excitation calculation procedure what cases to go through. Of course, the cases chosen depend on the desired mission of the particular deck. The procedure is as follows:

1. To calculate a case.
2. To check to see if all cases are done.
3. If all are done, stop.
4. If not, calculate the next case.
5. Go back to step 2.

In Appendix II, as mentioned earlier, a detailed discussion is presented to show how the logic circuitry of the computer is used to calculate field ampere-turns when a tapered air gap is used.

Program Utilization

The basic program just described is capable of calculating generator performance. However, there are two major hurdles which must be overcome in adapting the program for use in an engineering organization:

1. The engineer's time must not be wasted in translating data to or from computer language.
2. The program must have the versatility necessary to cover many different engineering jobs.

A digital computer is capable of absorbing large amounts of input data and emitting great quantities of output data. As a result, an immediate problem faced by a programmer is to arrange provisions for easy identification of the output data. The IBM *650* is capable of handling alphabetical information with output data identified as the output cards are punched. In this case an alternative approach was followed. The program is arranged so that punched output cards contain seven words (numbers) of data with a single numerical identifying number punched in the first word space on the card. Provision for identification of the output is made by typing alphabetical symbols identifying each item on separate cards corresponding to each card of output data. These cards, containing all of the heading information for the output, form an auxiliary deck which is used in

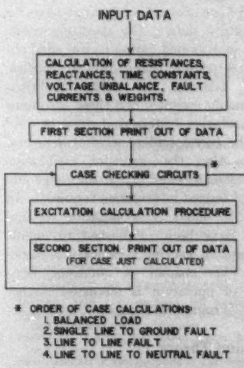

INPUT DATA
CALCULATION OF RESISTANCES, REACTANCES, TIME CONSTANTS, VOLTAGE UNBALANCE, FAULT CURRENTS & WEIGHTS.
FIRST SECTION PRINT OUT OF DATA
CASE CHECKING CIRCUITS
*
EXCITATION CALCULATION PROCEDURE
SECOND SECTION PRINT OUT OF DATA
(FOR CASE JUST CALCULATED)
STOP
* ORDER OF CASE CALCULATIONS
1. BALANCED LOAD
2. SINGLE LINE TO GROUND FAULT
3. LINE TO LINE FAULT
4. LINE TO LINE TO NEUTRAL FAULT

Fig. 5. Flow chart of generator design program

conjunction with the computer program itself. After the computer has produced the output, the output data cards and the heading card deck are assembled into a composite deck on an IBM type *77* collator, and the resulting composite deck is run through an IBM *407* accounting machine to provide a printed record of the output data with its accompanying identification. A sample output sheet is shown in Fig. 2. The sample also shows that provisions are made for several lines of identifying information to be printed ahead of the output.

As well as providing easily read output data, it is necessary to provide a simple means of specifying the input data. This is accomplished by the use of a preprinted input data form. This form is printed as a letter so that it may be sent by mail to the computer operator after the data are filled in. The form provides space for the necessary input data along with the symbols assigned to that item and a brief definition of each of the variables. Careful definition of each variable or constant on the input sheet greatly reduces the effort required of the new or infrequent user. The input data sheet also provides a brief key to the symbol lights displayed on the computer console when various program-error-sensing circuits are actuated. These displays simplify the tracing of errors when an invalid input prevents problem completion.

Provision for maximum use and versatility of the basic program has required its modification in a number of different ways. One of these modifications, as mentioned under "Program Input Data," was to establish a number of decks derived from the basic program. There were two objections to using one program for a wide range of problems.

1. An extremely large amount of output is obtained, at a considerable cost, from which the engineer is likely to receive more information than he needs or can reasonably use.

2. It is necessary to specify the full range of input information for each problem. This places an unnecessary burden on the engineer and lessens his interest in using the computer program.

The easiest way to minimize the time required to specify input data is to prepare from the basic program a number of auxiliary decks, each of which is designed to answer a rather small, specific problem. This approach is simple since it requires only that a copy of the basic deck be modified for the type of problem involved. The auxiliary deck can then be filed immediately adjacent to the computer and used whenever its type of problem is forwarded for solution.

A major factor in determiniug how the basic program is modified to establish auxiliary decks is the ease with which various input items can be changed. In general, machine dimensions and iron specifications are of primary importance; and in most cases they are changed with each problem. Temperatures and the range of output information, while they may change from problem to problem, are not as simple to modify in the basic program. Changing them requires considerably more effort. For this reason each auxiliary deck is set up for a fixed set of temperatures and a fixed set of output information and the engineer selects the appropriate deck by comparing the answers he needs to those provided by the various decks. The auxiliary decks now in use are listed in the following.

1. *Basic Program.* The basic program is maintained for use in the generation of new auxiliary decks. In addition, since this program has maximum flexibility, it is used for special studies of generator characteristics.

2. *Performance Analysis.* This auxiliary deck is a complete performance analysis program. For this deck, the engineer must specify all the machine dimensions, the iron to be used, and the generator rating. The outputs include design and complete performance data necessary for study of the generator, and also for evaluation of the complete electric system performance. The data are presented in a manner which permits prints of the output sheet to be forwarded directly to the interested control designers and system engineers. Separate performance analysis decks are maintained for the winding temperatures corresponding to each standard insulation system. The engineer selects the deck for the temperature range of the machine involved.

3. *Design Evaluation.* Also maintained in the computer file is a design evaluation deck. The input for this deck is the same as for the performance analysis deck, but the output is appreciably shortened. The output in this case includes only those machine constants and generator operating points useful to the design engineer in comparing similar generator designs.

4. *Application Data.* Another deck is provided which is useful not only to design engineers, but also to application engineers. In this deck, the input is only rating, operating conditions, and the model number of any standard production generator on which the design has already been fixed. Model number input is the key by which a set of punched cards is taken from a file listing complete machine dimensions for each production design. The computer operator inserts the particular set of cards as input on the basis of the model number specified on the problem input sheet. This type of deck makes for relatively simple analysis of off-rating operating conditions for various generators already in production.

5. *Off-Standard Design.* One other deck has been assembled which uses about half of the complete deck capability. Since the complete program is restricted in that it cannot handle unconventional types of generators, an auxiliary deck has been made up from the basic deck which eliminates the calculation of machine constants, but permits the calculation of excitation and other operating characteristics using machine constants as an input. This allows rapid and simple performance calculations on generators which the basic program cannot handle so far as the calculation of constants is concerned.

Table III. Test Over Calculated Results

Item	Test ÷ Calculated Ratios			
Generator rating (Kva)	20	22	40	60
Number of poles	6	4	8	8
Field current, balanced load				
A. Rated load	1.04	0.97	1.07	1.01
B. Double load	0.95	0.91	0.99	1.01
Six per-unit single line-to-ground fault				
A. Field current	1.02	0.97	0.99	1.04
B. Highest phase voltage	1.08	0.93	0.81	0.80
Negative-sequence reactance	0.99	1.07	0.91	1.00

Conclusions

The program described in this paper has proved capable of predicting generator performance within an acceptable degree of accuracy. Table III shows the relationship between test and calculated values of several of the program output quantities. The machines tabulated cover a fairly wide range in both speed and rating, giving some indication of the versatility of the program.

The program is also an extremely useful tool in areas other than the prediction of performance for normal types of machines. It is particularly useful in the analysis of the fundamental principles and techniques of design. With such a program it is relatively easy to introduce controlled variations in machine parameters and to observe the effect of these variations on geuerator characteristics and performance such as shown in Fig. 6. The effect of changes in rotor pole-face width (all other dimensions held constant)

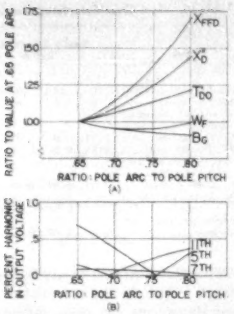


Fig. 6. Effect of variations in per-unit pole arc

A—Reactances and time constants
B—Harmonic voltages

on machine characteristics is shown in Fig. 6(A). Changes in pole arc have an important effect on the harmonic content of generator voltage. This is shown in Fig. 6(B). The data on harmonics were taken from another computer program, and included to provide a complete picture of the results of pole-arc variation. The relative magnitude of the harmonics shown were, of course, governed by the specific winding assumed for the study.

The availability of detailed generator characteristics on a wide variety of designs permits a much more effective analysis of new designs. An up-to-date design file of this type is a big help to the generator designer. It provides a comprehensive summary of design experience to aid in the development of new designs. One area in which the type of program described here proves to be of major value is in the analysis of the theories used in predicting machine performance. Specified characteristics of many different generators may be calculated according to any equation included in the program. This makes it relatively easy to establish the validity of new design techniques by trying out the equation on a large number of designs for which test data are available. Also, work now in process is utilizing probability theory to establish the amount of deviation which must be expected from normal or average data. It should then be possible to determine, on the basis of economic risk, the amount of margin which should be designed into machines of different types to meet specific customer requirements.

The major advantages gained by the use of the computer for aircraft generator performance calculations are:

1. The computer does a more rapid, accurate, and complete job of calculating generator performance than was previously possible.

2. Use of the computer frees the individual engineer from tedious, routine work and leaves him available for true creative engineering, which accelerates the pace of product improvement.

Appendix I. Balanced Load Excitation Calculation

Nomenclature

B_G = air-gap flux density
E = per-unit induced voltage per phase
F = air-gap ampere-turns at rated output voltage (unsaturated)
I = per-unit phase current
I_D = per-unit direct-axis phase current
I_Q = per-unit quadrature-axis phase current
K_1 = 1.0 plus stator iron mmf divided by air-gap mmf
M = total actual required field mmf
R_A = stator conductor resistance (per unit) per phase
R_F = field resistance (per unit referred to stator)
R_{KD} = amortisseur resistance in direct axis (per unit referred to stator)
R_{KQ} = amortisseur resistance in quadrature axis (per unit referred to stator)
R_2 = negative-sequence resistance (per unit referred to stator)
T_A = stator time constant
T_{DO}' = rotor (open-circuit stator) time constant
T_{DO}'' = amortisseur (open-circuit stator) time constant in direct axis
T_{QO}'' = amortisseur (open-circuit stator) time constant in quadrature axis
T_D' = direct-axis transient time constant
T_D'' = direct-axis subtransient time constant
T_Q'' = quadrature-axis subtransient time constant
V = per-unit terminal voltage per phase
W_F = field watts
X_A = stator leakage reactance (per unit) per phase
X_{AD} = stator synchronous reactance (per unit) in direct axis
$X_D = X_{AD} + X_A$
X_D' = transient reactance (per unit) in direct axis
X_D'' = subtransient reactance (per unit) in direct axis
X_{fld} = field leakage reactance (per unit)
$X_Q = X_{AQ} + X_A$
X_Q'' = subtransient reactance (per unit) in quadrature axis
X_0 = zero-sequence reactance (per unit)
X_2 = negative-sequence reactance (per unit)
X_{AQ} = stator synchronous reactance (per unit) in quadrature axis
α, see Fig. 1.
γ, See Fig. 1
θ = power-factor angle

With reference to Fig. 1, the induced voltage is

$$E = \sqrt{(V\cos\theta + IR_A)^2 + (V\sin\theta + IX_A)^2}$$

and the angles are

$$\alpha = \tan^{-1}\left(\frac{V\sin\theta + IX_A}{V\cos\theta + R_A}\right)$$

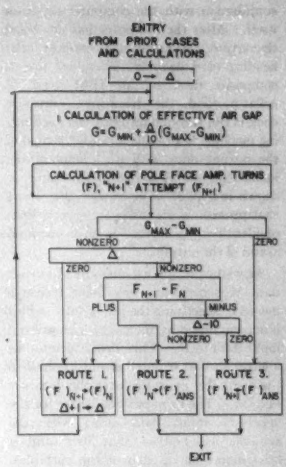


Fig. 7. Flow chart for obtaining excitation mmf including a tapered air gap

$$\gamma = \tan^{-1}\left(\frac{IX_{AQ} + K_1E\sin\alpha}{K_1E\cos\alpha}\right)$$

Then, the field mmf required to obtain terminal voltage V is

$$M = F(K_1E\cos(\gamma-\alpha) + I_DX_{AD}) + \text{rotor saturation mmf}$$

Note that

$$I_D = I\sin\gamma$$

The rotor saturation mmf is found by:

1. Calculating flux per pole as $E\cos(\gamma-\alpha)$ per unit and relating this to actual flux per pole by multiplying $E\cos(\gamma-\alpha)$ by the flux per pole necessary to produce rated voltage on open circuit.

2. Calculating rotor pole leakage factor as a leakage permeance factor times the known part of M plus one.

3. Multiplying item 1 by 2 and dividing by pole area to obtain pole flux density.

4. Determining rotor pole saturation mmf by entering the rotor iron saturation table using the flux density of item 3.

In this way, the total field excitation mmf M is now known for balanced loads of lagging power factor.

Appendix II. Example of Logic Circuitry for Tapered-Air-Gap Case

For generality and flexibility it was desired that the program handle the case of a tapered air gap over the poles. In this instance, a step-by-step method is needed to seek the solution for minimum pole-face ampere-turns since the "effective" air-gap length increases gradually from the mini-

mum to the maximum as saturation effects increase. Fig. 7 shows how the step-by-step method works. Jumps in gap of 10% from minimum to maximum are made. If no taper in the gap is present, *route 3* is followed. Following the path from entry, it is seen that, first, delta (Δ) is made zero and second, gap is calculated (in this case $G=G_{min}$). Then, pole-face ampere-turns are found, knowing G. A check of G_{max} minus G_{min} next gives zero, because G_{max} equals G_{min} when no taper in gap is present. This zero causes the path down *route 3* to be followed. By means of a choice circuit within the computer (which distinguishes between a zero number and a nonzero number) the proper path to follow is obtained.

When the machine has a tapered gap, the foregoing procedure is used to calculate pole-face ampere-turns, but now G_{max} minus G_{min} is a nonzero number. Thus, the computer proceeds into the Δ box, and with Δ still zero, exits down *route 1*. This route feeds a 1 (plus Δ, which is still zero) into the Δ box and so a new gap is calculated as G_{min} plus one tenth of the difference between maximum and minimum gaps. This gives a new value of pole-face ampere-turns (an Nth+1 value). The computer now proceeds into the Δ box where Δ is now also a nonzero number and so the next step is to subtract the original value of pole-face ampere-turns (F_n) from the new value (F_{n+1}). If this quantity is plus, i.e., the latest value of F is larger than the earlier value, then an exit through *Route 2* is taken and the answer is F_n. This is because the correct answer is the minimum number of ampere-turns.

If F_{n+1} is less than F_n, an exit through either *Route 1* or *3* is made from the (F_{n+1}-F_n) box. *Route 3* is followed if the value of Δ has been carried out to ten. In this case, the value of G has been increased all the way to G_{max}, because less ampere-turns were calculated as necessary each time δ was increased by one, and since at this point the gap is as large as possible, the correct answer is that for $G=G_{max}$. The path to *Route 1* from the (Δ-10) box is followed if gap has not been increased to G_{max} and yet the latest value of ampere-turns is less than the previous. In this case, Δ is increased by one and a recalculation of pole-face ampere-turns is made.

References

1. A-C Aircraft Generators. *Military Specification MIL-G-6099A.*

2. Calculation of Armature Reactance of Synchronous Machines, P. L. Alger. *AIEE Transactions*, vol. 47, Apr. 1928, pp. 493–513.

3. Calculation of Synchronous Machine Constants, L. Kilgore. *Ibid.*, vol. 50, Dec. 1931, pp. 1201–14.

4. The Direct- and Quadrature-Axis Equivalent Circuits of the Synchronous Machine, A. W. Rankin. *Ibid.*, vol. 64, Dec. 1945, pp. 851–68.

5. Circuit Analysis of A-C Power Systems (book), E. Clarke. John Wiley & Sons, Inc., New York, N. Y., vol. I, 1943.

6. Calculations on Voltage Unbalance for 3-Phase Synchronous Systems, B. J. Wilson, W. K. Gardner. *AIEE Transactions*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 426–37.

7. The Effect of Machine Impedances on the Voltage Unbalance of 3-Phase Synchronous Generators, T. F. Hartman. *Ibid.*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 467–71.

Staggered Sampling to Improve Stability of Multiple-Sampler Feedback Systems

RICHARD E. ANDEEN
ASSOCIATE MEMBER AIEE

THE PURPOSE of this paper is to present a technique which may be applied to improve the stability of sampled-data feedback control systems having more than one sampler without the use of additional compensating elements, and to illustrate by means of some examples the type of analysis procedure which may be used and the kind of results which may be expected. The technique which is referred to deals with the phasing of samplers in a multiple sampler feedback system to obtain an optimum condition of stability for a given configuration of the system. The technique is principally applicable either to existing systems where it is desired to improve upon a marginally satisfactory degree of stability in the most economical fashion, or to systems where the sampling results from inherent characteristics of the system data measuring or processing elements and is not introduced purposely to utilize the advantages of digital compensation. For the purposes of this paper the discussion will be limited to consideration of multiple sampler feedback systems containing only two samplers operating at the same sampling frequency and it will be assumed that a zero-order hold circuit follows each sampler. The results can be readily extended to systems with more than two samplers. These conditions are not exceedingly restrictive and many practical systems can be conceived which meet them in an engineering sense.

Paper 58-770, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted January 3, 1958; made available for printing March 21, 1958.

Richard E. Andeen is with the Sperry-Phoenix Company, Phoenix, Ariz.

The author wishes to express his thanks to Professor Richard W. Jones of Northwestern University for his encouragement and helpful suggestions, and also to acknowledge the financial support of Northwestern University and the Allstate Foundation's Barker Fellowship.

Phased Samplers

First, an attempt will be made to give a qualitative idea of the effect of staggered sampling. Fig. 1 represents a feedback system with two samplers in the forward path. For simplicity, the elements G_1 and G_2 are taken as simple lag networks with transfer functions $G_1=G_2=1/TS+1$. The total gain of the system is lumped in the amplifier K. Zero-order hold circuits follow each of the samplers. First, assume that the samplers are synchronized, that is, they both operate at $t=0$, T, $2T$, $3T$, ... Fig. 2 shows the relations between the various system voltages following the application of a 1-volt step input at $t=0$. The important thing to note in Fig. 2 is that two sampling periods are required, following the application of the step input, before the input E_3 to the first system element G_1 changes due to the feedback. There is thus a delay of two sampling periods in the transmission of a voltage change around the feedback loop.

Fig. 3 shows the relations between the system voltages when the samplers are phased so that S_1 operates at $t=0$, T, $2T$, $3T$, . . . and S_2 operates at $t=0.5T$, $1.5T$, $2.5T$, ... It is important to note here that, due to the staggering of the samplers, only one sampling period is required before the input E_3 to G_1 changes due to the feedback. The equivalent delay in transmission has been reduced to one sampling period. Intuitively it is felt that the second configuration of the system, all other things remaining equal, would be the more stable configuration

Fig. 1. Feedback system with two samplers

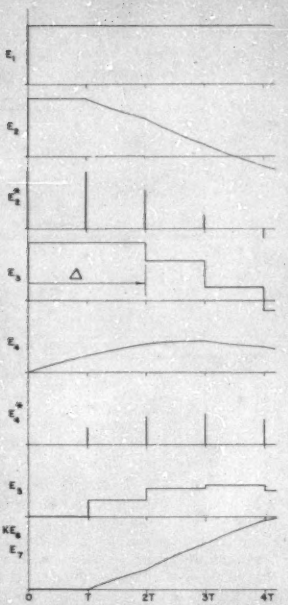


Fig. 2. Voltage relations with synchronized sampling: aT=1/3, K=4

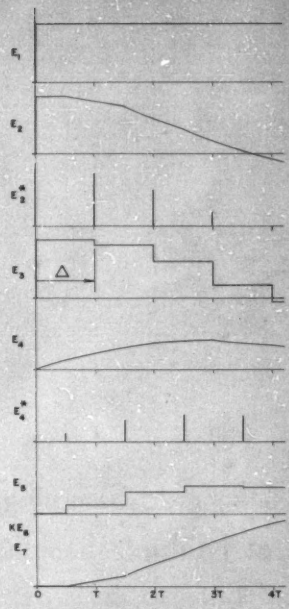


Fig. 3. Voltage relations with staggered sampling: aT=1/3, K=4

because of the reduced delay. This supposition is correct, as the mathematical analysis of this system will show in the next section.

If the samplers are phased by a time interval hT, this qualitative picture seems to indicate that an optimum value of h exists for which the stability of the system is a maximum. The samplers are in phase for $h=0$ and 1, so somewhere in between one would expect to find an optimum value. That this will depend upon the dynamic characteristics of the system and the sampling period T is obvious. The determination of the optimum value of h will also be discussed further in the next section.

Analysis

This analysis of staggered sampling is based on the use of the z-transform. Because of the assumption of hold circuits following each sampler, it is not necessary to consider the pulse width. Also, in this case it is known that the response of the system is well behaved between sampling instants. Thus the stability at the sampling instants is determinative.

By means of a suitable mathematical transformation of the system characteristic equationthe Routh-Hurwitz stability criterion can be applied and a stability boundary be investigated as is frequently done with continuous systems. The phasing of the samplers can be handled mathematically as shown in Fig. 4. A fictitious variable lead element is introduced before the phased sampler S_2 and an equal variable delay element introduced following it. It is then assumed that both samplers operate synchronously. The fictitious elements take care of the phasing.

First derive θ_1, which is the z-transform of the continuous element G_1 plus the first hold circuit and a fictitious lead element, and θ_2, which is the z-transform of the continuous element G_2 plus the second hold circuit and the fictitious delay element.

$$\theta_1(S)=HG_1(S)\epsilon^{hTS}=\left[\left(\frac{1-\epsilon^{-ST}}{S}\right)\left(\frac{a}{S+a}\right)\times\epsilon^{hTS}\right]$$

The z-transform can be evaluated by means of the method described by Jury.[1]

$$\theta_1(z)=\frac{1-z^{-1}}{2\pi i}\int_{c-j\infty}^{c+j\infty}\left(\frac{ae^{hTp}}{p(p+a)}\right)\times\left(\frac{dp}{(1-e^{-Ts+Tp})}\right)$$

Evaluating by residues,

$$\theta_1(z)=(1-z^{-1})\left[\frac{1}{1-z^{-1}}-\frac{e^{-ahT}}{1-e^{-aT}z^{-1}}\right] \quad (1)$$

Similarly,

$$\theta_2(S)=HG_2(S)e^{-hTS}=\left[\left(\frac{1-\epsilon^{-ST}}{S}\right)\left(\frac{a}{S+a}\right)\epsilon^{-hTS}\right]$$

$$\theta_2(z)=\frac{(1-z^{-1})z^{-1}}{2\pi i}\int_{c-j\infty}^{c+j\infty}\left(\frac{ae^{T(1-h)p}}{p(p+a)}\right)\times\left(\frac{dp}{(1-e^{-Ts+Tp})}\right)$$

$$\theta_2(z)=(1-z^{-1})z^{-1}\left[\frac{1}{1-z^{-1}}-\frac{e^{-aT(1-h)}}{1-e^{-aT}z^{-1}}\right] \quad (2)$$

Next, the characteristic equation of the system is formed,

$$1+K\theta_1(z)\theta_2(z)=0$$

$$1+Kz^{-1}(1-z^{-1})^2\left[\frac{1}{1-z^{-1}}-\frac{e^{-aTh}}{1-z^{-1}e^{-aT}}\right]\times\left[\frac{1}{1-z^{-1}}-\frac{e^{-aT(1-h)}}{1-z^{-1}e^{-aT}}\right]=0 \quad (3)$$

In order to apply the Routh-Hurwitz stability criterion the transformation $z=1+W/1-W$ must be made. After simplification the characteristic equation expressed in terms of W assumes the form,

$$a_0W^3+a_1W^2+a_2W+a_3=0 \quad (4)$$

where

$$a_0=D^2-KBC$$

$$a_1=D^2+2AD+KBC-KAB-KAC$$

$$a_2=2AD+KAB+KAC-KA^2+A^2$$

$$a_3=A^2(1+K)$$

$$A=1-e^{-aT}$$

$$B=1-2e^{-aTh}+e^{-aT}$$

$$C=1-2e^{-aT(1-h)}+e^{-aT}$$

$$D=1+e^{-aT}$$

For stability,

$$a_0>0,\ a_1>0$$

$$a_1a_2-a_0a_3>0$$

The last condition is determinative in this case; thus for the stability boundary,

$$[D^2+2AD+KBC-KAB-KAC]\times[2AD+KAB+KAC-KA^2+A^2]-[D^2-KBC][A^2(1+K)]=0 \quad (5)$$

Fig. 4. Feedback system with fictitious elements

Expanding this equation the following is obtained:

$$\gamma K^3+\beta K+\delta=0 \qquad (6)$$

where

$$\gamma=[(BC-AB-AC)(B+C-A)+ABC]$$

$$\beta=[(2D+A)(BC-AB-AC)+(D^2+2AD)(B+C-A)-AD^2+ABC]$$

$$\delta=[(D^2+2AD)(2D+A)-AD^2]$$

Solving for K,

$$K=\frac{-\beta\pm\sqrt{\beta^2-4\gamma\delta}}{2\gamma} \qquad (7)$$

Equation 7 gives the value of K at the stability boundary as a function of the variable h and the system parameters. Theoretically the optimum value of h could be determined by differentiating equation 7 with respect to h, setting dK/dh equal to zero, and solving for the value of h which yields a maximum value of K. Even in this very simple case, however, this procedure quickly leads to very lengthy mathematical expressions and tedious calculations. A simpler way of finding the optimum value of h is to solve equation 7 numerically for given values of h covering the range 0 to 1 and obtain the optimum value from a plot of the points obtained. That procedure was followed in this case; the results are shown in Fig. 5(A) for a particular choice of system parameters.

It is interesting to note in Fig. 5(A) that the optimum value of h occurs at $h=0.5$ and that the curve is symmetrical about this value. For $h=0.5$ a 25% increase in allowable gain at the stability boundary is obtained compared to the gain at $h=0$ or 1. A more dramatic improvement is illustrated in Fig. 5(B) which shows the results of a similar analysis, where the time constants of G_1 and G_2 are not equal but stand in the ratio $T_2/T_1=9$. A 100% increase in allowable gain is obtained in this case for a value of h equal to approximately 0.37.

The value of h obtained in the foregoing is optimum in the sense that it yields the greatest gain at the stability boundary and consequently yields minimum steady-state error. However, even for a system which is purely regulatory, it is obvious that operation at the stability boundary would not be consistent with satisfactory transient response. A reduced gain, all other things remaining constant, would be necessary in order to obtain an acceptable degree of damping in the transient response. It is intuitively appealing to believe that the optimum value of h will, for a given reduced gain, also yield the most stable response. In order to confirm this in a particular case one can determine the over-all system z-transform and calculate the transient response to a test input for several values of h in the vicinity of the so-called optimum value. In the case considered here, for example, it is known that the open-loop z-transform of the system is $K\theta_1(z)\theta_2(z)$. The calculation of transient responses by means of the z-transform technique would thus entail no unusual difficulties.

Graphical Solution

If the system under consideration is nonlinear, graphical procedure will often yield a solution rapidly. In this section a problem involving staggered sampling and a nonlinear system is worked out graphically. A hypothetical problem in the field of inventory control will be considered here in order to demonstrate the application of staggered sampling to a nonelectrical problem. The problem is represented in block diagram form in Fig. 6. It concerns the maintenance of a certain quantity of parts in a storage location from which parts are withdrawn for use in an adjacent assembly line. The parts are manufactured at a distant plant and shipped to the storage location. It is assumed that the parts are small parts such as machine screws and it is not economical to ship them from the manufacturing plant to the assembly plant more frequently than once a month. The cost of shipment depends principally on the labor and clerical work involved, and is almost independent of the number of parts shipped. Likewise, it is not economical to change the rate of manufacturing more than once a month because of setup costs.

The system originally established for controlling the parts inventory is as follows. On the first of the month the number of parts in the storage location is determined and compared to the number which is desired. This information is telegraphed to the manufacturing plant. At this point a nonlinearity is introduced. A decision is made to manufacture parts if the storage level is low. No parts are manufactured if the level is high. The manufacturing process is set up to manufacture parts at a uniform rate of production, the rate established so that the total quantity produced during the following month will equal the amount by which the storage level is low at the beginning of the month. In other words, $K_1=1/T$, where T is the sampling period, one month. At the end of the month the total quantity of parts which have accumulated is shipped to the assembly plant. Another nonlinearity is introduced at this point since the output of the manufacturing plant, θ_1, is reset to zero when the shipment occurs. Shipping requires 2 weeks, and thus represents a delay in the system of $\Delta=2$ weeks. It is assumed that the assembly line uses up parts at a constant rate, therefore the level of parts in the storage location drops continuously during the month.

Fig. 7 records the time history of the levels of parts in the storage location, the orders to manufacturing, and the level of manufacturing output, assuming that the storage location is initially empty. Fig. 7 was constructed very rapidly by a step-by-step graphical procedure and displays the important features of the system behavior. First, a large overshoot in the storage level is experienced, and then the storage level settles down to a steady-

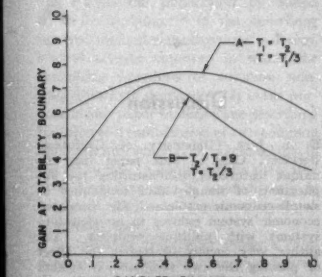


Fig. 5. Gain at stability boundary versus sampler phasing

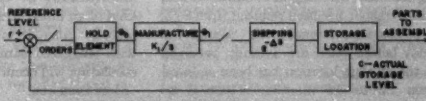


Fig. 6. Block diagram of inventory control problem

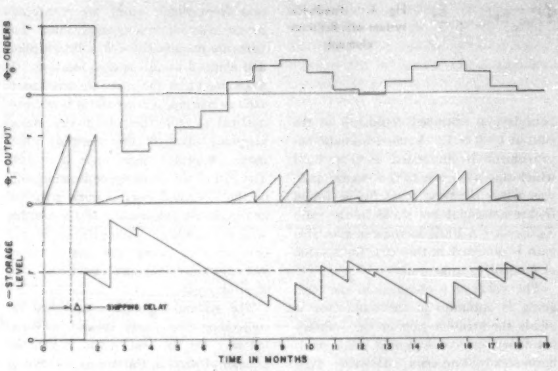


Fig. 7. Transient response of inventory control, original configuration

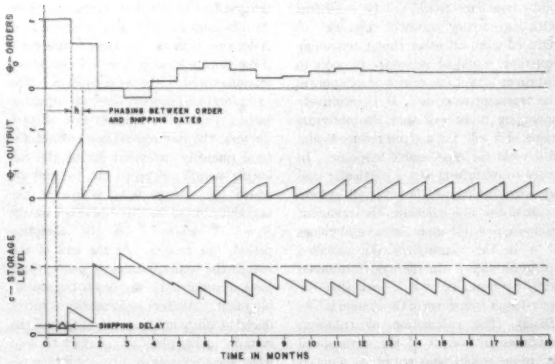


Fig. 8. Transient response of inventory control, with staggered sampling

state oscillation with a 6-month period. The rate of manufacturing activity, which is proportional to the slope of the θ_1 curve, varies considerably during the period of the oscillation necessitating the incurrence of unnecessary setup costs.

Fig. 8 illustrates the improvement which is obtained by staggered sampling. The configuration shown in Fig. 8 is identical to that in Fig. 7 with the one exception that the manufactured parts which have accumulated are shipped on the 12th of each month. It is evident from Fig. 8 that a considerable improvement has been obtained by simply phasing the order and shipment dates. First, the overshoot has been reduced from 80% to 20% of the reference level. Second, the delay before parts are initially received in the storage location has been reduced from 1.5 to 0.9 months. There is no steady-state oscillation and consequently the manufacturing setup costs are reduced in the long run.

This simple graphical analysis has brought out the salient characteristics of this hypothetical inventory control problem and indicates the phasing which will produce improved control and more economical production. Further reflection on this problem, stimulated by the graphical analysis, points out further interesting features of this system. In both Figs. 7 and 8 a steady-state error is maintained between the desired and actual storage levels. This perhaps can be corrected automatically by the provision of some kind of an integrating factor to the orders placed. Also, it is apparent that even with staggered sampling, a steady-state oscillation will occur if K_1 is either greater or less than one.

Conclusions

This paper has illustrated mathematical and graphical means for analyzing staggered sampling, and has presented the results of some examples. It has shown that the stability of multiple-sampler feedback control systems can be improved by the simple expedient of phasing the samplers. It should be apparent that the greater the number of samplers present in a system, the greater are the possibilities for improvement. This paper has considered only the improvement due to staggered sampling, but this is not to be construed as excluding the additional use of other stabilization means. Staggered sampling is a means for getting the most out of a given system configuration.

In determining the optimum degree of phasing, it is probably easier to calculate the system response for a number of values of h in the range 0 to 1 rather than to attempt more sophisticated mathematical methods. In cases involving more than two samplers the relative phasing of each is important, so that there will be values of $h_1, h_2, \ldots h_n$ to vary. In this case, if possible, it would be most favorable to set up the calculations so that they may be accomplished by digital computer means.

References

1. Analysis and Synthesis of Sampled-Data Control Systems, Eliahu I. Jury. *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

2. The Analysis of Sampled-Data Systems, J. R. Ragazzini, L. A. Zadeh. *Ibid.*, vol. 71, pt. II, Nov. 1952, pp. 225–34.

3. Extension of Continuous-Data System Design Techniques to Sampled-Data Control Systems, G. W. Johnson, D. P. Lindorff, C. G. A. Nordling. *Ibid.*, vol. 74, pt. II, Sept. 1955, pp. 252–63.

4. Synthesis and Critical Study of Sampled-Data Control Systems, Eliahu I. Jury. *Ibid.*, vol. 75, pt. II, July 1956, pp. 141–51.

5. Sampled-Data Control Systems, R. E. Andeen. *Report no. 2*, Northwestern University, Chicago, Ill., June 1957, pp. 24–33.

Discussion

E. I. Jury (University of California, Berkeley, Calif.): This paper has the unique feature of demonstrating the applications of sampled-data techniques to simple economic problems. The simulated economic system reduces to sampled-data systems with multiple samplers. The author very lucidly illustrates the effect of phasing the two samplers on system performance.

In this discussion, it is intended to emphasize that there exist many problems in economics,[1] production processes, and

operation research that can be solved using the sampled-data techniques. Usually, a simple economic problem in terms of its variables can be formulated as a linear difference equation with constant and periodic coefficients.[2] By effecting the z-transform technique these equations are reduced into operational form which can be represented then by a sampled-data configuration. Having obtained the sampled-data system analog of the economic system, then one can make use of the wealth of information obtained in this field during the last few years for the analysis, synthesis, and optimization of such systems. If the economic problem is nonlinear the graphical methods can be employed as the author has done in the paper.

The author has assumed the phasing of the samplers to be fixed at a certain desired value based on the system transfer function, however this parameter (phasing) could be varied as a function of the error and its derivatives. This would tend to make the system nonlinear; however the usefulness of such procedures should not be overlooked.

In concluding, the author is commended for introducing this new compensating procedure. It should be noted that this method can also be extended to higher order systems. Another method for system improvement is to have the second sampler, if possible, of a different period than the first.

REFERENCES

1. APPLICATIONS OF DISCRETE VARIABLE SERVO THEORY TO INVENTORY CONTROL. Herbert J. Vassian. *Journal*, Operations Research Society of America, Baltimore, Md., vol. 3, no. 3, Aug. 1955.

2. A NOTE ON THE OPERATIONAL SOLUTION OF DIFFERENCE EQUATIONS. E. I. Jury, F. J. Mullin. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 266, no. 3, Sept. 1958, p. 189.

Julius Tou (Purdue University, Lafayette, Ind.): This paper presents an interesting way of stabilizing a sampled-data feedback system which contains more than one sampler operating at the same rate but not synchronized. It so happens that this technique of stabilization has been taught in a sampled-data control course at Purdue University and has also been covered in a forthcoming book. It is probably worth while to point out that the author's analysis of the problem of staggered sampling (or sampled-data systems with nonsynchronized samplers as called by this discusser) could be simplified, especially when the controlled system is fairly complicated, if use is made of the modified z-transforms.

As can be seen, the z-transform associated with the transfer function $G(s)e^{-kTs}$ can readily be obtained from the modified z-transform associated with $G(s)$ by the substitution of $m=1-h$; that is,

$$Z\left\{G(s)e^{-kTs}\right\}=G(z,m)\Big|_{m=1-h}=G(z,1-h) \tag{8}$$

In like manner, it has been shown that the z-transform associated with $G(s)e^{kTs}$ is given by

$$\begin{aligned} Z\left\{G(s)e^{kTs}\right\} &= Z\left\{e^{Ts}G(s)e^{-(1-h)Ts}\right\} \\ &= zG(z,m)\Big|_{m=h} \\ &= zG(z,h) \end{aligned} \tag{9}$$

Since a fairly extensive table of the modified z-transforms is available in the classical paper of Barker,[1] to determine the z-transforms associated with $G(s)e^{-kTs}$ and $G(s)e^{kTs}$ is a simple matter.

The z-transform of the output of the error-sampled system shown in Fig. 1 of the paper, or its equivalent as shown in Fig. 2, is then given by the following general expression:

$$C(z)=\frac{Kz\overline{HG_1}(z,1-h)\overline{HG_2}(z,h)}{1+Kz\overline{HG_1}(z,1-h)\overline{HG_2}(z,h)}\,R(z) \tag{10}$$

The system characteristic equation is simply

$$1+Kz\overline{HG_1}(z,1-h)\overline{HG_2}(z,h)=0 \tag{11}$$

The output sequence of the system in response to an input is resulted from the inverse transform of equation 10. Equation 11 or its equivalent expression in w obtained through a bilinear transformation leads to an expression relating the factor h to the maximum allowable gain constant for stability, from which the most desirable value of h is computed. It is noted that when h is zero, i.e., these two samplers are synchronized, equation 11 reduces to the familiar expression

$$1+K\overline{HG_1}(z)\overline{HG_2}(z)=0 \tag{12}$$

REFERENCE

1. THE PULSE TRANSFER FUNCTION AND ITS APPLICATION TO SAMPLING SERVO SYSTEMS. R. H. Barker. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. IV, 1952, pp. 302–17.

The Describing Function of Backlash Followed by a Dead Zone

ALFRED D. GRONNER
NONMEMBER AIEE

AFTER the publication by Ralph Kochenburger[1,2] of the describing function method of nonlinear servo design, several articles appeared in which the describing functions of common nonlinearities were derived.[2,3] It is the purpose of this paper to derive the describing function of a combination of two common nonlinearities which is not yet available in the literature. Graphs will be presented in a form that can be used directly in nonlinear feedback system analysis as commonly used.

In addition, this paper shows a comparison of the describing function derived directly for the combined nonlinearity with the approximate describing function derived by considering the describing functions of backlash and dead zone in cascade.

Derivation of Describing Function

The nonlinearity to be treated consists of backlash H (see Fig. 1) followed by a dead zone D (see Fig. 2). This combination appears, for example, in a servo system with backlash on the output gear train and dead zone in the amplifier, if the inertia after the backlash can be neglected.

Refer to Fig. 3, which shows the response of a backlash element to a sinusoidal input function $u=U\sin\alpha$. If a dead zone $\pm D$ follows this element, its output Y will be zero for $|X|<D$ and reduced by $\pm D$ for $|X|<D$. If the following definitions are used, together with the ones shown in Figs. 1, 2, and 3, the equation for the normalized output of the dead-zone element, Z, is obtained:

$$A=\frac{H}{U} \qquad \sin\beta=1-A$$

$$B=\frac{D}{H} \qquad \sin\gamma=(B-1/2)A$$

$$Z=\frac{Y}{H} \qquad \sin\delta=(B+1/2)A$$

$$Z=\frac{1}{A}=\frac{1}{2}-B \qquad \text{for } \frac{\pi}{2}<\alpha<\pi-\beta$$

$$Z=\frac{1}{A}\sin\alpha+\frac{1}{2}-B \qquad \text{for } \pi-\beta<\alpha<\pi-\gamma$$

$$Z=0 \qquad \text{for } \pi-\gamma<\alpha<\pi+\delta$$

$$Z=\frac{1}{A}\sin\alpha+\frac{1}{2}+B \qquad \text{for } \pi+\delta<\alpha<\frac{3\pi}{2} \tag{1}$$

Paper 58-771, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted December 12, 1957; made available for printing March 21, 1958.

ALFRED D. GRONNER is with the American Machine and Foundry Company, Greenwich, Conn.

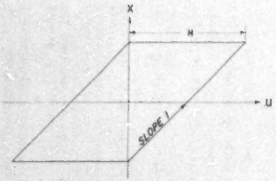


Fig. 1. Backlash element

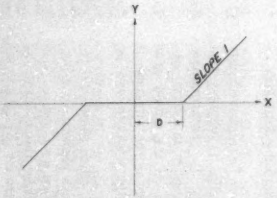


Fig. 2. Dead-zone element

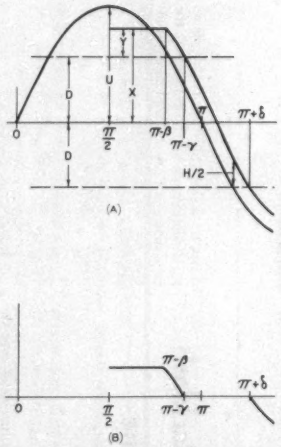


Fig. 3. Output of backlash element (A) and dead-zone element (B)

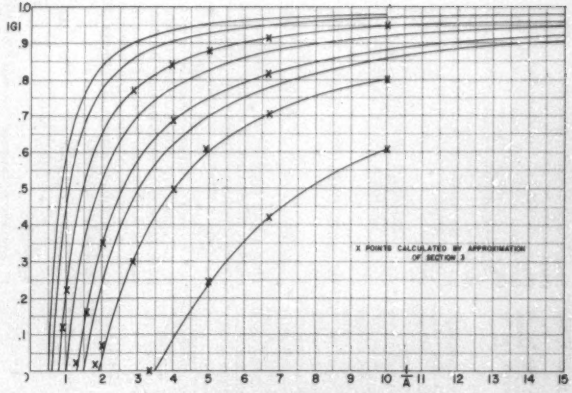


Fig. 4. Absolute value of describing function

Table I

				B = 0.3				B = 0.8				B = 1.4				B = 3.0			
1/A	A	G_H	φ	D/V	$\lvert G_D \rvert$	|G|	$\lvert G_{DB} \rvert$	D/V	$\lvert G_D \rvert$	|G|	$\lvert G_{DB} \rvert$	D/V	$\lvert G_D \rvert$	|G|	$\lvert G_{DB} \rvert$	D/V	$\lvert G_D \rvert$	|G|	$\lvert G_{DB} \rvert$
0.5	1.8	0.125	65.3																
0.71	1.4	0.368	46.6																
0.77	1.3	0.426	42.9	0.915	0.03	0.013	37.7												
0.83	1.2	0.483	39.3	0.745	0.152	0.073	22.7												
1.00	1.0	0.593	32.5	0.506	0.385	0.228	12.9												
1.25	0.80	0.697	26.0					0.918	0.03	0.021	34.6								
1.54	0.65	0.771	21.2	0.253	0.685	0.527	5.67	0.675	0.213	0.164	15.7								
1.82	0.55	0.817	18.1									0.942	0.018	0.015	37.0				
2.00	0.5	0.839	16.5	0.179	0.775	0.650	3.8	0.476	0.420	0.352	9.1	0.834	0.079	0.066	23.6				
2.85	0.35	0.901	11.7	0.116	0.855	0.770	2.3	0.316	0.606	0.545	5.3	0.553	0.335	0.302	9.5				
3.33	0.30	0.920	10.1													0.977	0.006	0.005	45.2
4.0	0.25	0.938	8.5	0.080	0.900	0.844	1.5	0.213	0.733	0.686	3.3	0.373	0.537	0.504	6.0	0.800	0.104	0.097	20.2
5.0	0.20	0.955	6.9	0.0629	0.922	0.880	1.1	0.167	0.790	0.754	2.5	0.283	0.650	0.620	4.2	0.629	0.257	0.247	12.2
6.67	0.15	0.970	5.2	0.0463	0.945	0.918	0.7	0.124	0.845	0.820	1.7	0.216	0.730	0.707	3.0	0.464	0.433	0.420	7.5
10.0	0.10	0.983	3.5	0.0306	0.963	0.945	0.5	0.0815	0.898	0.884	1.1	0.143	0.820	0.805	1.9	0.306	0.618	0.609	4.3

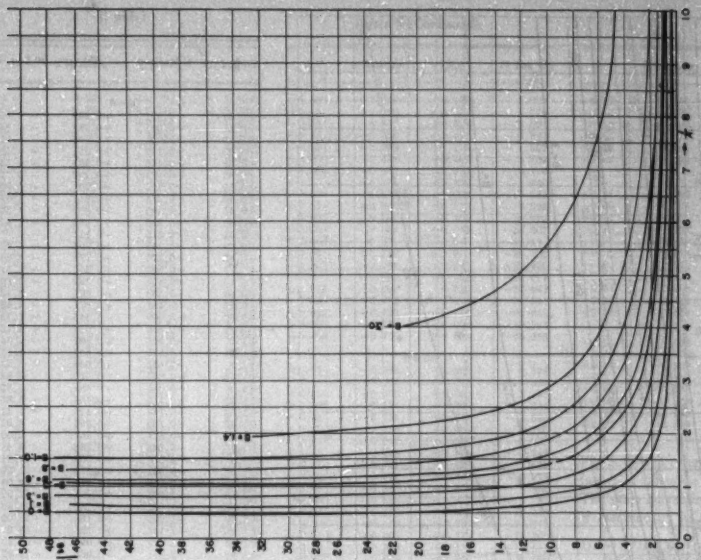
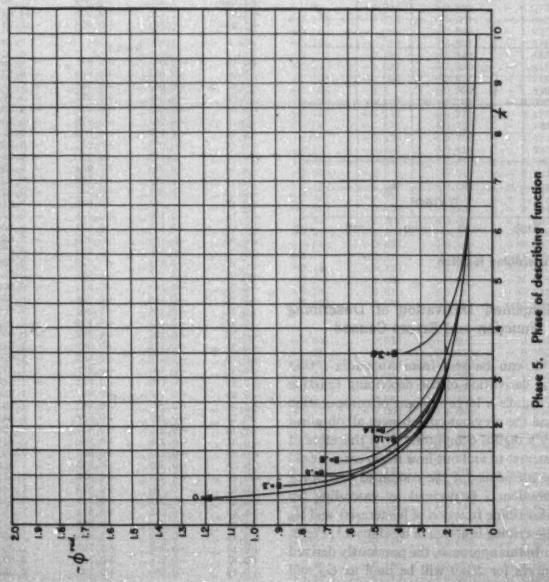


Fig. 6. Absolute value of describing function in decibels



Phase 5. Phase of describing function

As the describing function of a nonlinearity is defined as the complex ratio of the fundamental of the output over the (sinusoidal) input, the fundamental of $Z(\alpha)$ must be obtained first. If a_1 is called the fundamental component in quadrature with the input, and b_1 the fundamental component in phase with the input, then

$$a_1 = \frac{2}{\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} Z \cos a \, da\alpha$$

$$b_1 = \frac{2}{\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} Z \sin a \, da\alpha \qquad (2)$$

(a_1, b_1 are normalized with respect to H.)

The describing function G is then (by definition)

$$G = \frac{\sqrt{a_1^2+b_1^2}}{\frac{1}{A}} \exp\left(j \tan^{-1}\frac{a_1}{b_1}\right) =$$

$$|G| \exp(j\phi) \quad (3)$$

$$|G| = A\sqrt{a_1^2+b_1^2} \quad \phi = \tan^{-1}\frac{a_1}{b_1}$$

$$|G|^{dB} = 20 \log_{10} A\sqrt{a_1^2+b_1^2}$$

From equations 1, 2, and 3 the describing function is derived in Appendix I. A tabulation of $|G|$, ϕ $20 \log_{10}|G|$, with A

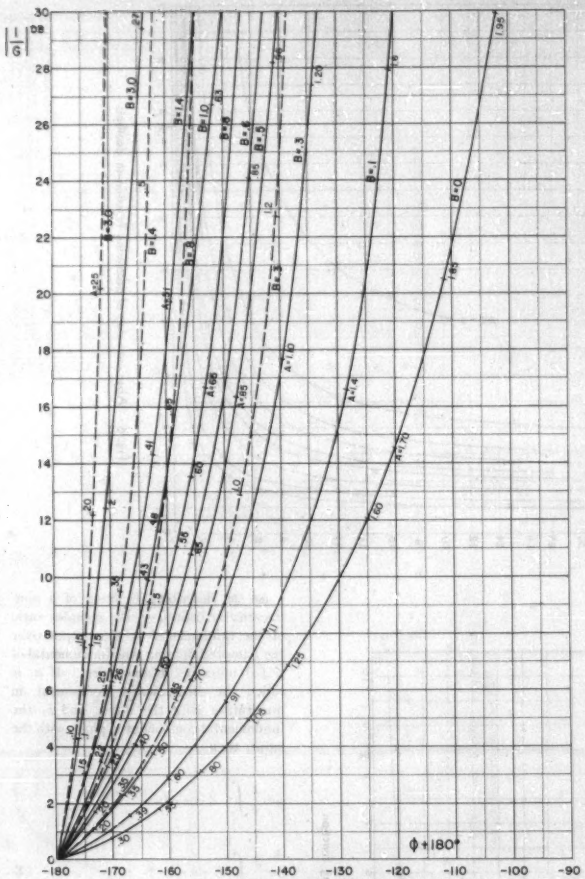


Fig. 7. Nichols plot of describing function

and B as parameters, is given in Table I.

As it was the intention of this paper to present the results in a form directly applicable to the describing function method (for description of this method see, for example, reference 4), the following graphs are presented: Fig. 4—$|G|$ versus $1/A$; Fig. 5—ϕ versus $1/A$; Fig. 6—20 $\log_{10}|G|$ versus $1/A$; Fig. 7—$-20\log_{10}|G|$ versus $-\phi$. In all cases, B is given as the parameter (see Table II).

In the opinion of the author, Fig. 7 is the most useful one, as it can be used in the method of analysis shown in reference 4, page 578, without any replotting.

Simplified Derivation of Describing Function and Errors Caused

It can be seen from Appendix I that the derivation of the describing function G entails a large amount of computation and the previous curves were all obtained by a digital computer. It is therefore of interest to find out how much error would be introduced if the combined describing function G is derived by cascading G_H (describing function of hysteresis) and G_D (describing function of dead space); Fig. 8.

In this approach, the previously derived curves for $B=0$ will be used as G_H and

Table II

A	ϕ, Radians	$\lvert G\rvert$	20 Log $\lvert G\rvert$
	B=0		
+0.05	−0.031	+0.993	− 0.05
+0.1	−0.061	+0.983	− 0.14
+0.15	−0.091	+0.969	− 0.26
+0.20	−0.120	+0.954	− 0.40
+0.25	−0.148	+0.938	− 0.55
+0.30	−0.173	+0.920	− 0.72
+0.35	−0.205	+0.901	− 0.90
+0.40	−0.233	+0.814	− 1.09
+0.45	−0.260	+0.860	− 1.30
+0.50	−0.288	+0.839	− 1.52
+0.55	−0.315	+0.816	− 1.75
+0.60	−0.343	+0.794	− 2.00
+0.65	−0.370	+0.770	− 2.26
+0.70	−0.398	+0.745	− 2.53
+0.75	−0.426	+0.722	− 2.82
+0.80	−0.453	+0.697	− 3.13
+0.85	−0.481	+0.671	− 3.45
+0.90	−0.509	+0.645	− 3.79
+0.95	−0.538	+0.619	− 4.16
+1.0	−0.566	+0.592	− 4.54
+1.05	−0.595	+0.565	− 4.94
+1.1	−0.625	+0.538	− 5.37
+1.15	−0.655	+0.510	− 5.83
+1.2	−0.685	+0.482	− 6.32
+1.25	−0.716	+0.454	− 6.85
+1.30	−0.748	+0.425	− 7.41
+1.35	−0.780	+0.396	− 8.02
+1.40	−0.814	+0.367	− 8.69
+1.45	−0.848	+0.338	− 9.41
+1.5	−0.884	+0.308	−10.21
+1.55	−0.921	+0.278	−11.09
+1.60	−0.960	+0.248	−12.09
+1.65	−1.001	+0.218	−13.22
+1.70	−1.045	+0.187	−14.55
+1.75	−1.092	+0.156	−16.09
+1.80	−1.144	+0.125	−18.0
+1.85	−1.202	−0.094	−20.47
+1.90	−1.271	+0.063	−23.97
+1.95	−1.359	+0.031	−29.98
+1.955	−1.370	+0.028	−30.88
+1.96	−1.381	+0.025	−31.90
+1.965	−1.394	+0.022	−33.05
+1.97	−1.407	+0.019	−34.39
+1.975	−1.421	+0.015	−35.97
+1.98	−1.437	+0.012	−37.91
+1.985	−1.455	+0.009	−40.40
+1.99	−1.476	+0.006	−43.92
+1.995	−1.504	+0.003	−49.94
	B=0.1		
+0.05	−0.031	+0.987	− 0.109
+0.06	−0.037	+0.984	− 0.137
+0.07	−0.043	+0.981	− 0.166
+0.08	−0.049	+0.977	− 0.196
+0.09	−0.055	+0.974	− 0.228
+0.10	−0.061	+0.970	− 0.260
+0.11	−0.067	+0.966	− 0.294
+0.12	−0.073	+0.962	− 0.328
+0.13	−0.079	+0.958	− 0.364
+0.14	−0.085	+0.954	− 0.400
+0.15	−0.091	+0.950	− 0.437
+0.16	−0.097	+0.946	− 0.475
+0.17	−0.103	+0.942	− 0.514
+0.18	−0.109	+0.938	− 0.554
+0.19	−0.115	+0.933	− 0.594
+0.20	−0.120	+0.929	− 0.635
+0.21	−0.126	+0.924	− 0.677
+0.22	−0.133	+0.920	− 0.720
+0.23	−0.138	+0.915	− 0.764
+0.24	−0.144	+0.911	− 0.808
+0.25	−0.149	+0.906	− 0.853
+0.26	−0.155	+0.901	− 0.899
+0.27	−0.161	+0.896	− 0.945
+0.28	−0.167	+0.892	− 0.992
+0.29	−0.172	+0.887	− 1.040
+0.30	−0.178	+0.882	− 1.088
+0.31	−0.184	+0.877	− 1.137
+0.32	−0.189	+0.872	− 1.187
+0.33	−0.195	+0.867	− 1.238
+0.34	−0.201	+0.862	− 1.289
+0.35	−0.206	+0.856	− 1.341
+0.36	−0.212	+0.851	− 1.394
+0.37	−0.218	+0.846	− 1.447
+0.38	−0.223	+0.841	− 1.501
+0.39	−0.229	+0.835	− 1.556
+0.40	−0.235	+0.830	− 1.612
+0.41	−0.240	+0.825	− 1.668
+0.42	−0.246	+0.819	− 1.725
+0.43	−0.252	+0.814	− 1.783
+0.44	−0.258	+0.808	− 1.841
+0.45	−0.263	+0.803	− 1.900
+0.46	−0.268	+0.797	− 1.900
+0.47	−0.274	+0.792	− 2.021
+0.48	−0.280	+0.786	− 2.082

Table II (Continued)

A	ϕ Radians	$\lvert G\rvert$	20 Log $\lvert G\rvert$
+0.49	−0.285	+0.781	− 2.144
+0.50	−0.291	+0.775	− 2.207
+0.51	−0.297	+0.769	− 2.271
+0.52	−0.302	+0.764	− 2.335
+0.53	−0.308	+0.758	− 2.400
+0.54	−0.313	+0.752	− 2.466
+0.55	−0.319	+0.746	− 2.533
+0.56	−0.325	+0.741	− 2.601
+0.57	−0.330	+0.735	− 2.669
+0.58	−0.336	+0.729	− 2.739
+0.59	−0.341	+0.723	− 2.809
+0.60	−0.347	+0.717	− 2.880
+0.61	−0.353	+0.711	− 2.952
+0.62	−0.358	+0.705	− 3.024
+0.63	−0.364	+0.699	− 3.098
+0.64	−0.370	+0.693	− 3.173
+0.65	−0.375	+0.687	− 3.248
+0.66	−0.381	+0.681	− 3.325
+0.71	−0.400	+0.651	− 3.723
+0.76	−0.438	+0.620	− 4.145
+0.81	−0.466	+0.589	− 4.596
+0.86	−0.495	+0.557	− 5.079
+0.91	−0.524	+0.534	− 5.592
+0.96	−0.554	+0.492	− 6.156
+1.0	−0.584	+0.459	− 6.760
+1.0	−0.614	+0.425	− 7.416
+1.1	−0.646	+0.392	− 8.132
+1.1	−0.677	+0.358	− 8.922
+1.2	−0.710	+0.323	− 9.797
+1.2	−0.743	+0.289	−10.780
+1.3	−0.778	+0.254	−11.898
+1.3	−0.814	+0.216	−13.190
+1.4	−0.852	+0.183	−14.720
+1.4	−0.892	+0.148	−16.590
+1.5	−0.934	+0.112	−18.990
+1.6	−1.036	+0.040	−27.954
+1.6	−1.117	+0.004	−46.553
	B = 0.3		
+0.05	−0.031	+0.974	− 0.222
+0.07	−0.043	+0.963	− 0.325
+0.08	−0.049	+0.957	− 0.379
+0.09	−0.055	+0.951	− 0.434
+0.1	−0.062	+0.944	− 0.491
+0.11	−0.068	+0.938	− 0.549
+0.12	−0.074	+0.932	− 0.608
+0.13	−0.080	+0.925	− 0.669
+0.14	−0.086	+0.919	− 0.730
+0.15	−0.092	+0.912	− 0.793
+0.20	−0.122	+0.878	− 1.124
+0.25	−0.151	+0.842	− 1.484
+0.30	−0.181	+0.805	− 1.873
+0.35	−0.210	+0.768	− 2.292
+0.40	−0.239	+0.729	− 2.743
+0.45	−0.269	+0.689	− 3.229
+0.50	−0.298	+0.649	− 3.753
+0.55	−0.328	+0.608	− 4.320
+0.60	−0.358	+0.566	− 4.936
+0.65	−0.388	+0.524	− 5.607
+0.70	−0.419	+0.481	− 6.343
+0.75	−0.450	+0.438	− 7.155
−0.80	−0.482	+0.395	− 8.059
+0.85	−0.514	+0.351	− 9.076
+0.90	−0.548	+0.307	−10.236
+0.95	−0.583	+0.263	−11.581
+1.0	−0.619	+0.219	−13.180
+1.05	−0.658	+0.174	−15.145
+1.10	−0.699	+0.130	−17.686
+1.15	−0.745	+0.086	−21.278
+1.20	−0.801	+0.042	−27.421
+1.205	−0.807	+0.038	−28.354
+1.21	−0.814	+0.033	−29.397
+1.215	−0.821	+0.029	−30.579
+1.22	−0.828	+0.025	−31.942
+1.225	−0.836	+0.021	−33.553
+1.23	−0.845	+0.016	−35.523
+1.235	−0.855	+0.012	−38.060
+1.24	−0.866	+0.008	−41.630
+1.245	−0.880	+0.004	−47.718
	B = 0.5		
+0.05	−0.031	+0.961	− 0.336
+0.06	−0.037	+0.953	− 0.411
+0.07	−0.043	+0.945	− 0.487
+0.08	−0.05	+0.936	− 0.566
+0.09	−0.056	+0.928	− 0.646
+0.1	−0.062	+0.919	− 0.728
+0.11	−0.068	+0.910	− 0.812
+0.12	−0.074	+0.901	− 0.897
+0.13	−0.080	+0.892	− 0.984
+0.14	−0.086	+0.883	− 1.073
+0.15	−0.092	+0.874	− 1.164
+0.20	−0.123	+0.827	− 1.641
+0.25	−0.153	+0.779	− 2.162
+0.30	−0.184	+0.730	− 2.731
+0.35	−0.214	+0.679	− 3.351

A	ϕ Radians	$\lvert G\rvert$	20 Log $\lvert G\rvert$
+0.40	−0.245	+0.628	− 4.029
+0.45	−0.276	+0.577	− 4.776
+0.50	−0.308	+0.524	− 5.601
+0.55	−0.340	+0.471	− 6.521
+0.60	−0.373	+0.418	− 7.557
+0.65	−0.407	+0.365	− 8.739
+0.70	−0.442	+0.312	−10.110
+0.75	−0.479	+0.258	−11.740
+0.80	−0.518	+0.205	−13.742
+0.85	−0.560	+0.152	−16.329
+0.90	−0.608	+0.100	−19.979
+0.95	−0.666	+0.048	−26.307
+0.955	−0.672	+0.043	−27.151
+0.96	−0.679	+0.038	−28.205
+0.965	−0.684	+0.033	−29.400
+0.97	−0.695	+0.028	−30.777
+0.975	−0.703	+0.023	−32.404
+0.98	−0.713	+0.019	−34.391
+0.985	−0.723	+0.014	−36.949
+0.99	−0.735	+0.009	−40.543
+0.995	−0.750	+0.004	−46.663
	B = 0.6		
+0.05	−0.031	+0.955	− 0.394
+0.1	−0.062	+0.906	− 0.849
+0.15	−0.093	+0.855	− 1.355
+0.20	−0.124	+0.802	− 1.911
+0.25	−0.154	+0.748	− 2.521
+0.30	−0.185	+0.692	− 3.190
+0.35	−0.217	+0.636	− 3.927
+0.40	−0.248	+0.579	− 4.743
+0.45	−0.281	+0.521	− 5.652
+0.50	−0.313	+0.463	− 6.673
+0.55	−0.347	+0.405	− 7.837
+0.60	−0.383	+0.347	− 9.182
+0.65	−0.419	+0.289	−10.773
+0.70	−0.458	+0.231	−12.713
+0.75	−0.501	+0.173	−15.192
+0.80	−0.548	+0.117	−18.615
+0.85	−0.604	+0.061	−24.166
+0.90	−0.690	+0.008	−40.918
	B = 0.8		
+0.05	−0.031	+0.942	− 0.510
+0.06	−0.037	+0.930	− 0.622
+0.07	−0.044	+0.918	− 0.736
+0.08	−0.05	+0.906	− 0.853
+0.09	−0.056	+0.893	− 0.973
+0.1	−0.062	+0.881	− 1.096
+0.11	−0.069	+0.868	− 1.221
+0.12	−0.075	+0.856	− 1.349
+0.13	−0.081	+0.843	− 1.480
+0.14	−0.087	+0.830	− 1.613
+0.15	−0.094	+0.817	− 1.749
+0.16	−0.1	+0.804	− 1.888
+0.17	−0.106	+0.791	− 2.030
+0.18	−0.113	+0.778	− 2.175
+0.19	−0.119	+0.765	− 2.323
+0.20	−0.125	+0.752	− 2.474
+0.25	−0.157	+0.685	− 3.280
+0.30	−0.189	+0.617	− 4.180
+0.35	−0.222	+0.549	− 5.193
+0.40	−0.256	+0.481	− 6.344
+0.45	−0.291	+0.413	− 7.671
+0.50	−0.328	+0.345	− 9.234
+0.55	−0.367	+0.277	−11.135
+0.60	−0.409	+0.211	−13.516
+0.65	−0.456	+0.145	−16.743
+0.70	−0.512	+0.081	−21.786
+0.75	−0.590	+0.021	−33.370
+0.755	−0.601	+0.015	−36.076
+0.76	−0.615	+0.01	−39.948
+0.765	−0.633	+0.004	−46.882
	B = 1.0		
+0.05	−0.031	+0.930	− 0.628
+0.06	−0.037	+0.915	− 0.765
+0.07	−0.044	+0.900	− 0.906
+0.08	−0.05	+0.886	− 1.050
+0.09	−0.056	+0.871	− 1.198
+0.1	−0.063	+0.856	− 1.349
+0.11	−0.069	+0.840	− 1.504
+0.12	−0.075	+0.825	− 1.663
+0.13	−0.082	+0.810	− 1.825
+0.14	−0.088	+0.795	− 1.991
+0.15	−0.095	+0.779	− 2.161
+0.16	−0.101	+0.764	− 2.335
+0.17	−0.107	+0.748	− 2.513
+0.18	−0.114	+0.733	− 2.695
+0.19	−0.120	+0.717	− 2.882
+0.20	−0.127	+0.701	− 3.073
+0.21	−0.133	+0.686	− 3.269
+0.22	−0.140	+0.670	− 3.470
+0.23	−0.146	+0.654	− 3.675
+0.28	−0.180	+0.576	− 4.730
+0.33	−0.215	+0.497	− 6.009

A	ϕ Radians	$\lvert G\rvert$	20 Log $\lvert G\rvert$
+0.38	−0.251	+0.418	− 7.556
+0.43	−0.289	+0.340	− 9.353
+0.48	−0.330	+0.263	−11.583
+0.53	−0.376	+0.188	−14.506
+0.58	−0.429	+0.116	−18.780
+0.63	−0.498	+0.046	−26.711
	B = 1.4		
+0.01	−0.006	+0.991	− 0.161
+0.06	−0.038	+0.885	− 1.059
+0.11	−0.07	+0.785	− 2.067
+0.16	−0.103	+0.684	− 3.294
+0.21	−0.132	+0.582	− 4.687
+0.26	−0.174	+0.482	− 6.336
+0.31	−0.213	+0.382	− 8.341
+0.36	−0.256	+0.285	−10.880
+0.41	−0.304	+0.192	−14.327
+0.46	−0.364	+0.103	−19.906
+0.51	−0.453	+0.022	−32.782
	B = 3.0		
+0.05	−0.032	+0.803	− 1.899
+0.1	−0.068	+0.607	− 4.335
+0.15	−0.108	+0.417	− 7.592
+0.2	−0.159	+0.240	−12.378
+0.25	−0.233	+0.086	−21.309
+0.26	−0.255	+0.059	−24.577
+0.27	−0.282	+0.033	−29.388
+0.28	−0.322	+0.011	−38.985

Fig. 8.1-11 of reference 5 as G_D. To obtain G_D, calculate D/V for each input. The same normalized input ($1/A = U/H$) is assumed as in the preceding. Then:

$$1/A = U/H;\ 1/A\,|G_N| = V/H;\ V/B_N = V/D;$$
$$D/V = AB/|G_N|;\ |G| = |G_N|\,|G_D|;$$
$$\angle G = \angle G_N \quad (4)$$

$|G|$ is calculated for $B = 0.3$, 0.8, 1.4, 3.0 from Table I, and the points are indicated on Figs. 4 and 7 by crosses and dashes.

It can be seen that although the input to the dead-zone nonlinearity is not at all sinusoidal, the error in the magnitude of G is quite small.

As the dead zone has no phase shift, the phase of G as computed by this method would only depend on A and not on B. From Fig. 5 this is only true for small A's (input over four times H). For smaller inputs, the phase shift of the approximate describing function is significantly smaller than without approximation. The magnitude of the error can best be seen on Fig. 7, where the approximate describing functions of $B = 0.3$, 0.8, 1.4, 3.0, are shown by dashed lines.

Conclusions

The describing function of backlash followed by dead zone has been derived, and presented in easily usable graphs.

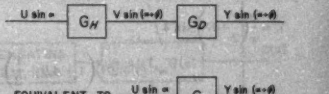


Fig. 8. Block diagram for approximate describing function

Furthermore, the errors caused by cascading transfer functions instead of deriving the transfer function for the combined nonlinearity have been evaluated. It is hoped that the graphs presented will be of help in stabilizing backlash oscillations.

Appendix I. Derivation of Describing Function

1. Evaluation of a_1:

$$\frac{2}{\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} Z\cos\alpha\, d\alpha=\frac{2}{\pi}\int_{\frac{\pi}{2}}^{\pi-\beta}\left(\frac{1}{A}-\frac{1}{2}-B\right)\times\cos\alpha\, d\alpha+\frac{2}{\pi}\int_{\pi-\beta}^{\pi-\gamma}\frac{1}{A}\sin\alpha\cos\alpha\, d\alpha+\frac{2}{\pi}\int_{\pi-\beta}^{\pi-\gamma}\left(\frac{1}{2}-B\right)\cos\alpha d\delta+\frac{2}{\pi}\int_{\pi+\delta}^{\frac{3\pi}{2}}\frac{1}{A}\sin\delta\cos\delta d\delta+\frac{2}{\pi}\int_{\pi+\delta}^{\frac{3\pi}{2}}\left(\frac{1}{2}+B\right)\cos\delta d\delta$$

$$=a_1=I_1+I_2+I_3+I_4+I_5$$

$$I_1=\frac{2}{\pi}\left(\frac{1}{A}-\frac{1}{2}-B\right)\sin\alpha\Big/_{\frac{\pi}{2}}^{\pi-\beta}$$

$$=\frac{2}{\pi}\left(\frac{1}{A}-\frac{1}{2}-B\right)(1-A-1)=\frac{1}{\pi}(A+2AB-2)$$

$$I_2=\frac{1}{A\pi}\frac{\cos 2\delta}{2}\Big/_{\pi-\beta}^{\pi-\delta}=\frac{-1}{2\pi A}(\cos 2\delta-\cos 2\gamma)$$

$$=\frac{-1}{2\pi A}[1-2\sin^2\beta-(1-2\sin^2\gamma)]$$

$$=\frac{1}{2\pi A}\left[1-2\left(B-\frac{1}{2}\right)^2A^2-1+2(1-A)^2\right]$$

$$=\frac{1}{A\pi}[(1-A)^2-A^2\left(B-\frac{1}{2}\right)^2$$

$$I_3=\left(\frac{1}{2}-B\right)\frac{2}{\pi}\sin\alpha\Big/_{\pi-\beta}^{\pi-\gamma}=\frac{2}{\pi}\left(\frac{1}{2}-B\right)\frac{2}{\pi}\times\left[\left(B-\frac{1}{2}\right)A-1+A\right]=\frac{1}{\pi}(1-2B)\times\left(AB+\frac{1}{2}-1\right)$$

$$I_4=\frac{-1}{A\pi}\cos 2\alpha\Big/_{\pi+\delta}^{\frac{3\pi}{2}}=\frac{1}{2\pi A}(1-\cos 2\gamma)$$

$$=\frac{1}{2\pi A}(1-1+2\sin^2\delta)=\frac{1}{A\pi}A^2\left(B+\frac{1}{2}\right)^2$$

$$I_5=\frac{2}{\pi}\left(\frac{1}{2}+B\right)\sin\delta\Big/_{\pi+\delta}^{\pi-\beta}$$

$$=\frac{1}{\pi}(1+2B)\left(1+AB+\frac{A}{2}\right)$$

Multiplying out and simplifying,

$$a_1=I_1+I_2+I_3+I_4+I_5=\frac{1}{\pi}(A+2AB-2)\quad(5)$$

2. Evaluation of b_1:

$$b_1=\frac{2}{\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} Z\sin\delta\, d\delta=\frac{2}{\pi}\int_{\frac{\pi}{2}}^{\pi-\beta}\left(\frac{1}{A}-\frac{1}{2}-B\right)\sin\delta d\delta+\frac{2}{\pi}\int_{\pi-\beta}^{\pi-\gamma}\frac{1}{A}\sin^2\delta d\delta+\frac{2}{\pi}\int_{\pi-\beta}^{\pi-\gamma}\left(\frac{1}{2}-B\right)\sin\delta d\delta+\frac{2}{\pi}\int_{\pi+\delta}^{\frac{3\pi}{2}}\frac{1}{A}\sin^2\delta d\delta+\frac{2}{\pi}\int_{\pi+\delta}^{\frac{3\pi}{2}}\left(\frac{1}{2}+B\right)\sin\delta d\delta=J_1+J_2+J_3+J_4+J_5$$

$$J_1=\frac{2}{\pi}\left(\frac{1}{A}-\frac{1}{2}-B\right)\cos\beta$$

$$=\frac{1}{A\pi}(2-A-2AB)\cos\beta=J_1$$

$$J_2=\frac{2}{A\pi}\left[\frac{1}{2}\delta-\frac{1}{4}\sin 2\alpha\right]\Big/_{\pi-\beta}^{\pi-\gamma}$$

$$=\frac{1}{A\pi}\left[\beta-\gamma+\frac{1}{2}\sin 2\gamma-\frac{1}{2}\sin 2\beta=J_2\right.$$

$$J_3=\frac{2}{\pi}\left(B-\frac{1}{2}\right)\cos\delta\Big/_{\pi-\beta}^{\pi-\gamma}=\frac{1}{A\pi}(2AB-A)\times(\cos\beta-\cos\gamma)=\frac{1}{A\pi}(2AB-A)\times\cos\beta-2\sin\gamma\cos\gamma\Big]=\frac{1}{A\pi}[(2AB-A)\cos\beta-\sin 2\gamma]=J_3$$

$$J_4=\frac{2}{A\pi}\left[\frac{1}{2}\delta-\frac{1}{4}\sin 2\delta\right]\Big/_{\pi+\delta}^{\frac{3\pi}{2}}=\frac{1}{A\pi}\left[\frac{\pi}{2}-\delta+\frac{1}{2}\sin 2\delta\right]=J_4$$

$$J_5=\left(B+\frac{1}{2}\right)\frac{2}{\pi}\cos\delta\Big/_{\frac{3\pi}{2}}^{\pi+\delta}=\frac{1}{A\pi}\,2A\times\left(B+\frac{1}{2}\right)\cos\delta=\frac{1}{A\pi}\sin\delta\cos\delta$$

$$=\frac{1}{A\pi}\sin 2\delta=J_5$$

$$J_1+J_3=\frac{1}{A\pi}(2-2A-2AB-A)\cos\beta$$

$$=\frac{1}{A\pi}2(1-A)\cos\beta=\frac{1}{A\pi}\,2\times\sin\beta\cos\beta=\frac{1}{A\pi}\sin 2\beta$$

3. Adding all J's and collecting terms,

$$b_1=\frac{1}{A\pi}\left[\frac{\pi}{2}-\delta-\gamma+\beta+\frac{1}{2}\sin 2\beta-\frac{1}{2}\sin 2\gamma-\frac{1}{2}\sin 2\delta\right]$$

By using the foregoing equation, the describing function can be obtained.

Table III

B	T, K_{max}	DB, K_{max}	C, %*
0	3.1	9.8	0
0.1	5.4	14.6	10
0.3	10.0	20.9	30
0.5	15.5	23.8	50
1.0	23.7	27.5	100
1.4	47.3	33.5	140
3.0	168.0	44.5	300

* $C=100\,B$ increase in inaccuracy due to dead zone compared to backlash alone (in per cent).

Appendix II

Following the procedures of reference 6, the maximum stable gain (K_{max}) in a feedback loop containing an element of transfer function $K/S(Ts+1)$ followed by backlash+dead zone as described before can be obtained graphically. The improvement obtained can be seen in Table III.

References

1. Frequency Response Method for Analysing and Synthesizing Contactor Servomechanisms, **Ralph J. Kochenburger.** *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 270–84.

2. Limiting in Feedback Control Systems, **Ralph J. Kochenburger.** *Ibid.*, vol. 72, pt. II, July 1953, pp. 180–94.

3. Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities, **H. D. Greif.** *Ibid.*, Sept., pp. 243–48.

4. Automatic Feedback Control Synthesis (book), **J. G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955, chap. 10.

5. Servomechanisms and Regulating System Design, Vol. II (book), **Harold Chestnut, Robert W. Mayer.** John Wiley & Sons, Inc., New York, N. Y., 1955.

6. Backlash in a Velocity Lag Servomechanism, **Nathaniel B. Nichols.** *AIEE Transactions*, vol. 72, pt. II, 1953 (Jan. 1954 section), pp. 462–67.

Discussion

Otto J. M. Smith (University of California, Berkeley, Calif.): Mr. Gronner has done an excellent job of computing the amplitude-dependent gain of a combination of dead zones and backlash. In Fig. 7 he has a plot of a whole class of these amplitude-dependent transferences on a Nichols plot. When these curves are used in conjunction with a plot of the sinusoidal transference of the linear part of the system, then the Nichols plot becomes a conformal transformation of the Laplace s-plane and the describing function or the amplitude-dependent locus is the root locus in the s-plane. However, a Nichols plot is a left-handed co-ordinate system, and many of the valuable features of the root locus technique are lost by virtue of this reversal of the signs of all the angles. Furthermore, distances in the s-plane have significance concerning relative damping and resonant frequencies, but these distances have been distorted in the Nichols plot because the scale for the gain co-ordinate is different

from the scale for the phase co-ordinate when the conformal transformation has been taken.

I would like to recommend that all plots of gain versus phase be strictly rigorous conformal transformations of the root loci from the s-plane. This means that the scale of log gain should have units of nepers which are exactly equal in length to the units of radians on the scale of phase. Strictly speaking, the log gain is the log of the absolute magnitude of the gain function plus j times the phase of the gain function in radians. Therefore the co-ordinate system should be plotted in units such that every 8 decibels has the same length as 40 degrees in phase shift.

Since the log gain function is the log of the absolute magnitude of the gain plus j times the phase lead, if positive values of gain are up on the page then this is the real axis of this complex plane and therefore the j axis should be to the left and positive values of phase should be to the left.

A detailed discussion of this log gain plane and its characteristics is given in reference 1. The $B=0$ curve in Fig. 7 of Mr. Gronner's paper is plotted in reference 1, Fig. 13–21, page 458, on these co-ordinates of log gain. In addition, describing functions are given for wave shapes other than sine waves.

Mr. Gronner is to be highly complimented on his contribution to the library of analysis tools being made available to feedback control engineers.

REFERENCE

1. FEEDBACK CONTROL SYSTEMS (book), Otto J. M. Smith. McGraw-Hill Book Company, Inc., New York, N. Y., 1958, pp. 29–42.

Alfred D. Gronner: Professor Smith's suggestion is an excellent one and I am grateful for it, especially because I intend to publish other describing functions in the same form in a forthcoming book, and was looking for a standard format. I intend to use the log gain plane in the future.

The Design of Analog Computer Compensated Control Systems

STEPHEN C. BIGELOW
NONMEMBER AIEE

THE PROBLEM of designing a continuous feedback control system may generally be stated in the following form:

Given

1. The characteristics of the plant to be controlled, usually in terms of a transfer function.
2. A set of specifications of the desired characteristics of the over-all system. These specifications may cover such system attributes as transient response, frequency response, steady-state error to particular inputs, certain error constants, etc.

Determine

1. The over-all transfer function of a system, which meets the specifications.
2. The specifications of a compensating system, which in combination with the plant results in a system with the desired over-all transfer functions.

This paper presents a method of solving this problem for a particular class of control systems, which parallels that developed for the design of sampled data control systems.[1–3] The restrictions on the specifications of the over-all system imposed by the characteristics of the plant are stated explicitly.

Paper 58-800, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 10, 1958; made available for printing March 17, 1958.

STEPHEN C. BIGELOW is with Columbia University, New York, N. Y.

The class of control systems to be considered is that of tandem compensated duplicators. A block diagram of such a system is shown in Fig. 1. It may be readily shown that the design equation for the compensating system $D(s)$ is as follows:

$$D(s)=\frac{1}{G(s)}\,\frac{K(s)}{1-K(s)} \qquad (1)$$

where $G(s)$ is the plant transfer function and $K(s)$ is the over-all system transmission function, $C(s)/R(s)$. Since $G(s)$ is assumed to be given, $D(s)$ is completely specified by a knowledge of $K(s)$. It is further assumed, as a matter of convenience, that $D(s)$ is to be realized by means of an analog computer consisting only of summers and integrators. It is shown in Appendix I that $D(s)$ may be expressed as

$$D(s)=\frac{a_0s^n+a_1s^{n-1}+\ldots+a_{n-1}s+a_n}{s^n+b_1s^{n-1}+\ldots+b_{n-1}s+b_n} \qquad (2)$$

where the a's and b's are constants directly related to scaling factors in the analog computer realization of $D(s)$.

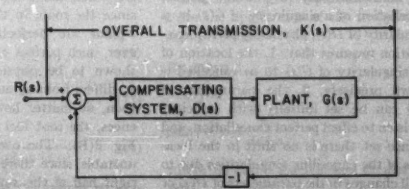


Fig. 1. Error-compensated duplicator control system

Restrictions on Specification of Over-all Transmission, K(s)

Certain restrictions are placed on the specifications of the over-all system transmission function $K(s)$ by stability considerations and the characteristics of the plant.

STABILITY

For the over-all system to be stable, the roots of the characteristic equation must all have negative real parts. Any convenient method may be used to determine whether or not this condition is met for a particular $K(s)$, e.g., the Routh-Hurwitz criterion, etc.

EXCESS OF POLES OF K(s) OVER ZEROS OF K(s)

If it is assumed that the plant, $G(s)$, has no poles outside the finite s-plane, then, writing $G(s)$ as the ratio of two rational polynomials in s, $P(s)$, and $Q(s)$, $G(s)=P(s)/Q(s)$, it follows that the order of $Q(s)$, q, must be greater than or equal to the order of $P(s)$, p. Therefore, in the finite s-plane the excess number of poles over zeros of $G(s)$ is $q-p$.

Since it is assumed that the compensating system $D(s)$ is to be realized by means of an analog computer without differentiators, $D(s)$ can have no poles

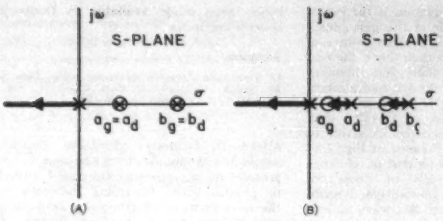


Fig. 2. System root loci illustrating perfect cancellation (A), and imperfect cancellation (B)

outside the finite s-plane. Writing $D(s) = A(s)/B(s)$ where $A(s)$ and $B(s)$ are rational polynomials in s, it is clear that the order of $B(s)$, b, must be greater than or equal to the order of $A(s)$, a.

The over-all transmission, $K(s)$ is related to $D(s)$ and $G(s)$ by

$$K(s) = \frac{D(s)G(s)}{1+D(s)G(s)} \tag{3}$$

Substituting $P(s)/Q(s)$ for $G(s)$ and $A(s)/B(s)$ for $D(s)$ equation 3 becomes

$$K(s) = \frac{A(s)P(s)}{B(s)Q(s)+A(s)P(s)} = \frac{U(s)}{V(s)} \tag{4}$$

where $U(s) = A(s)P(s)$ and $V(s) = B(s)Q(s)+A(s)P(s)$. Since the sum and product of rational polynomials are themselves rational, $U(s)$ and $V(s)$ are rational polynomials in s. The order of $U(s)$, u, is $a+p$. The order of $V(s)$, v, is the maximum of $a+p$ and $b+q$ which is $b+q$, since b is greater than or equal to a and q is greater than or equal to p. Since $v = b+q$ is greater than or equal to $u = a+p$, it is clear that $K(s)$ can have no poles outside the finite s-plane. Within the finite s-plane $K(s)$ must then have an excess number of poles over zeros equal to $v-u = b+q-(a+p) = (b-a)+(q-p)$. Therefore, the number of excess poles over zeros of $K(s)$ must exceed the number of excess poles over zeros of $G(s)$ by the number of excess poles over zeros of $D(s)$.

Cancellation of Poles and Zeros

As the result of certain physical considerations, restrictions must be placed on the cancellation of poles and zeros of the plant $G(s)$ by zeros and poles of the compensating system $D(s)$. While it is possible mathematically to obtain perfect cancellation of a singularity of $G(s)$ by a singularity of $D(s)$, physically perfect cancellation requires that: 1. the location of the singularity of $G(s)$ to be cancelled is known precisely, 2. the parameters of $D(s)$ can be set initially with sufficient precision to effect perfect cancellation, and 3. once set there is no shift in the locations of the cancelling singularities due to small changes in the parameters of $D(s)$ or $G(s)$. Clearly, these conditions cannot all be met. It is therefore physically impossible to cancel perfectly poles and zeros of the plant $G(s)$ with zeros and poles of the compensating system $D(s)$.

Consider then the effect on system performance of imperfect cancellation of a pair of singularities. So long as the real parts of these singularities remain less than zero, the system will be stable and will behave as though perfect cancellation had been effected at all frequencies except those in the neighborhood of the imperfectly cancelled pair of singularities. However, if the imperfectly cancelled pair of singularities have real parts greater than zero the system characteristic equation will have at least one root with a positive real part, and the system will be unstable. To illustrate this point consider a plant whose transfer function is $G(s) = (s-a_g)/s(s-b_g)$, where a_g and b_g are real and greater than zero. Then $G(s)$ has a zero at $s = a_g$ and poles at $s = 0$ and $s = b_g$ where the subscript g indicates that these singularities are physically in the plant $G(s)$. Suppose one were to attempt to cancel in the compensating system $D(s)$ the pole of $G(s)$ at $s = b_g$ and the zero of $G(s)$ at $s = a_g$. Then $D(s) = (s-b_d)/(s-a_d)$ where the subscript d indicates that these singularities are physically in $D(s)$. The root loci for the resulting over-all system as a function of some parameter, say plant gain k_g, may be determined from the expression

$$k_g D(s)G(s) = k_g \frac{s-b_d}{s-a_d} \frac{s-a_g}{s(s-b_g)} = -1$$

If $a_g = a_d$ and $b_g = b_d$, the characteristic equation root loci are as shown in Fig. 2(A) and the over-all system is stable, since the roots in the right half of the s-plane are perfectly cancelled. However, such perfect cancellation has been shown to be physically impossible. If a_g differs from a_d and b_g differs from b_d, then, no matter how small these differences, the root loci will be as shown in Fig. 2(B). The over-all system is then unstable, since there are two roots in the right half of the s-plane.

Thus it is seen that physical cancellation of poles or zeros of the plant $G(s)$, by zeros or poles of the compensating system $D(s)$, must be restricted to those poles and zeros which lie in the left half of the s-plane, sufficiently far from the $j\omega$-axis so that they always remain in the left half of the s-plane, even under the influence of normal variations in system parameters.

To avoid inadvertent violation of this restriction, it is convenient to write it in terms of restrictions on the over-all transmission $K(s)$. Noting that

$$K(s) = \frac{D(s)G(s)}{1+D(s)G(s)} \tag{5}$$

it is readily seen that any zero of $G(s)$ which is not cancelled by a pole of $D(s)$ must also be a zero of $K(s)$. Therefore, to ensure that a zero of $G(s)$ is not cancelled by a pole of $D(s)$, it is necessary that this zero of $G(s)$ be a zero of $K(s)$ also. Noting further that

$$1-K(s) = 1-\frac{D(s)G(s)}{1+D(s)G(s)} = \frac{1}{1+D(s)G(s)} \tag{6}$$

it is apparent that any pole of $G(s)$ which is not cancelled by a zero of $D(s)$ must also be a zero of $1-K(s)$. Therefore, to ensure that a pole of $G(s)$ is not cancelled by a zero of $D(s)$, it is necessary that this pole of $G(s)$ be a zero of $1-K(s)$ also.

Thus the restrictions on the cancellation of zeros and poles of $G(s)$ by poles and zeros of $D(s)$ may be stated as follows: All singularities of $G(s)$ whose real parts are greater than zero (or less than, but sufficiently near, zero so that it is possible for them to become greater than zero under normal system parameter variations) must be treated as follows: 1. all such zeros of $G(s)$ must be zeros of $K(s)$, and 2. all such poles of $G(s)$ must be zeros of $1-K(s)$.

Specification of Over-all System Transmission, K(s)

In the preceding section the restrictions on the over-all system transmission $K(s)$ have been stated. The first step in the solution of the design problem is to choose a system transmission function $K(s)$, which is consistent both with the foregoing restrictions and with the given specifications. In general there will be a large class of functions consistent with both of these conditions.

At this point perhaps a few remarks about the underlying philosophy of this design problem are in order. Systems of the type being considered are commonly

called duplicators, since ideally their purpose is to duplicate at the output the input time function. The given design specifications, no matter how stated, serve to impose on the final system minimum standards of performance in response to certain types of inputs. Usually, these inputs belong to the class of time functions $f_n(t)$, whose general form is

$$f_n(t) = t^n \mu(t) \tag{7}$$

where $n = 0, 1, 2, \ldots$, and $\mu(t) =$ unit step function. The response of duplicator control systems to such inputs may be broken up into two parts; the initial or transient response when time t is relatively small, and the steady-state response when t is very large. Usually, it is desired that during the transient part of the system response the output be brought into approximate correspondence with the input in some optimum fashion and that during the remainder of the response the difference between input and output be made as small as possible.

It would be desirable to determine explicitly the relationship between the system time response to inputs of the type described and the over-all system transmission function $K(s)$. In this connection it is convenient to consider system transient and steady-state response separately.

Steady-State Response

The steady-state performance of a feedback control system in response to inputs of the form $t^n\mu(t)$ is usually expressed in terms of the system steady-state error or in terms of the generalized error constants.

The system error $E(s)$ is defined as the difference between system input and output, i.e.,

$$E(s) = R(s) - C(s) = R(s)[1 - K(s)] \tag{8}$$

The steady-state system error to a particular input may be found by means of the final value theorem. For example, if the input is a unit step, $\mu(t)$, $R(s) = 1/s$ and the steady-state error $e(\infty)$, is

$$e(\infty) = \lim_{s \to 0} sE(s) = \lim_{s \to 0} [1 - K(s)]$$

For zero steady-state error, $1 - K(s)$ must have a zero at $s = 0$ or $1 - K(s)$ must contain s as a factor. This result is generalized in Table I.

These conditions on $1 - K(s)$ for zero steady-state error may be conveniently written in terms of conditions on the over-all transmission $K(s)$. Recalling that by equation 4, $K(s) = U(s)/V(s)$ where $U(s)$ is a rational polynomial in s of order u, and $V(s)$ is a rational polynomial in s of order v, let

$$U(s) = c_u s^u + c_{u-1} s^{u-1} + \ldots + c_1 s + c_0$$

and

$$V(s) = d_v s^v + d_{v-1} s^{v-1} + \ldots + d_1 s + d_0$$

where the c's and d's are constants and as previously shown v is greater than or equal to u. Then $1 - K(s)$ may be written

$$1 - K(s) = 1 - \frac{U(s)}{V(s)} = \frac{V(s) - U(s)}{V(s)}$$

$$= \frac{\begin{array}{l} d_v s^v + d_{v-1} s^{v-1} + \ldots + d_{u+1} s^{u+1} + \\ (d_u - c_u) s^u + (d_{u-1} - c_{u-1}) s^{u-1} + \\ \ldots + (d_1 - c_1) s + (d_0 - c_0) \end{array}}{V(s)} \tag{9}$$

For $1 - K(s)$ to contain as a factor s^{n+1}, it is clear from inspection of equation 9 that the following relations must all hold

$$\begin{aligned} d_0 - c_0 &= 0 \\ d_1 - c_1 &= 0 \\ &\vdots \\ d_n - c_n &= 0 \end{aligned} \tag{10}$$

If the relations of equation 10 are all satisfied, then the system will have zero steady-state error to inputs of the form $f_n(t) = t^n\mu(t)$.

The generalized system error constants[4,6] K_p, K_v, and K_a may be defined in terms of the coefficients of the Maclaurin series expansion of $E(s)/R(s)$

$$\frac{E(s)}{R(s)} = 1 - K(s)$$

$$= \frac{1}{1 + K_p} + \frac{1}{K_v} s + \frac{1}{K_a} s^2 + \ldots \tag{11}$$

It is readily shown that K_p, K_v, and K_a, the position, velocity, and acceleration error constants, are inverse measures of the steady-state system error in response to unit step, velocity, and acceleration inputs respectively.

Since control system design problems frequently involve obtaining specified error constants, it is convenient to relate these constants to the coefficients c_i and d_i of the over-all system transmission $K(s)$. The derivation of these relations is given in Appendix II. In summary the results are

$$K_p = \frac{c_0}{d_0 - c_0} \tag{12}$$

$$K_v = \frac{d_0^2}{c_0 d_1 - c_1 d_0} \tag{13}$$

$$K_a = \frac{d_0^3}{d_0(d_2 c_0 - d_0 c_2) - d_1(d_1 c_0 - d_0 c_1)} \tag{14}$$

Thus it is seen that the system steady-state response can be related explicitly to the over-all system transmission $K(s)$.

Table I. Conditions on $1 - K(s)$ for Zero Steady-State Error

Input	Factors Which Must Be Contained in $1 - K(s)$
Unit step $\mu(t)$	s
Unit ramp $t\mu(t)$	s^2
Unit acceleration $t^2\mu(t)$	s^3
In general for input $= t^n\mu(t)$	s^{n+1}

Transient Response

The specifications given in control system design problems often include such characteristics of system transient response as rise-time and peak overshoot in response to a unit step input. Even when no explicit specifications are given concerning transient response, it is generally assumed that the system transient response should be "good" in some accepted sense.

There are several well-known methods of choosing an over-all system transmission $K(s)$, which will give approximately the desired system transient response.[5-7] A particularly simple method to apply is that suggested by Graham and Lathrop.[7-9] This method involves the use of tables of standard forms of system transmissions, such as those published by these authors.

Graham and Lathrop have tabulated normalized system transmissions which give optimum transient response as judged by the ITAE (integral of time-multiplied absolute-value of error) criterion which minimizes the integral, I

$$I = \int_0^\infty t\,|\,\epsilon(t)\,|\,dt \tag{15}$$

Tables based on other criteria are also available.[10] While these standard forms of system transmission may not be consistent with all the given design specifications, usually only slight modifications are required to make them consistent.

Examples

The following examples serve to illustrate the application of the foregoing design procedure.

EXAMPLE 1

Given

plant $G(s) = \dfrac{K_x}{s(s+a)}$

where K_x and a are real and greater than zero.

Specifications

1. Bandwidth should be less than 100 rad/sec (radians per second).
2. K_p should be infinite.
3. K_v should be greater than 50/sec.

Solution

Step 1. Stability will be assured by the use of a standard form for $K(s)$, provided that such use does not result in cancellation of singularities of $G(s)$ which lie in or very near the right half of the s-plane.

Step 2. $G(s)$ has two excess poles over zeros in the finite s-plane. Therefore $K(s)$ must have at least two excess poles over zeros.

Step 3. $G(s)$ has no poles in the right half of the s-plane. However, it has a pole on the $j\omega$-axis at $s=0$. This pole should be a zero of $1-K(s)$.

Step 4. The lowest order ITAE standard form which has two excess poles over zeros in the finite s-plane is the second order zero position error form

$$K(s)=\frac{\omega_0^2}{s^2+1.4\omega_0 s+\omega_0^2}$$

and

$$1-K(s)=\frac{s(s+1.4\omega_0)}{s^2+1.4\omega_0 s+\omega_0^2}$$

It is noted that

a. $K(s)$ is stable, since it is a standard form and $1-K(s)$ has a zero at $s=0$.

b. $K(s)$ has two excess poles over zeros.

c. $K_p=\dfrac{c_0}{d_0-c_0}=\dfrac{\omega_0^2}{\omega_0^2-\omega_0^2}=\infty$

d. $K_v=\dfrac{d_0^2}{c_0d_1-c_1d_0}=\dfrac{\omega_0^4}{1.4\omega_0^3}=\dfrac{\omega_0}{1.4}$

Inspection of normalized frequency response curves for ITAE standard forms shows that system bandwidth is approximately equal to ω_0. Therefore, let $\omega_0=99$ rad/sec, which meets the specifications that bandwidth be less than 100 rad/sec. Then, the velocity constant $K_v=99/1.4=70$/sec, which meets the specifications that K_v be greater than 50/sec. It is seen that

$$K(s)=\frac{(99)^2}{s^2+1.4(99)s+(99)^2}=\frac{9{,}800}{s^2+139s+9{,}800}$$

meets all design requirements.

Step 5. By equation 1

$$D(s)=\frac{1}{G(s)}\frac{K(s)}{1-K(s)}=\frac{s(s+a)}{K_2}\frac{9{,}800}{s(s+139)}$$

$$=\frac{9{,}800}{K_2}\frac{s+a}{s+139}$$

Let $a_0=9{,}800/K_2$, $a_1=a/a_0$, and $b_1=139$. Then

$$D(s)=\frac{a_0s+a_1}{s+b_1}$$

The analog computer realization for $D(s)$ is shown in Fig. 3.

EXAMPLE 2

Consider the problem of example 1 with the following single change in specifications: Bandwidth should be less than 50 rad/sec. The solution is the same as before through step 4d

$$K_v=\frac{\omega_0}{1.4}$$

Let $\omega_0=49$ rad/sec, which meets the new bandwidth specifications. However, $K_v=49/1.4=34$, which is less than the specified minimum of 50/sec. Therefore the specifications cannot be met with the second-order zero position error ITAE standard form of $K(s)$. The next logical step is to try a higher order standard form in the hope that enough degrees of freedom will be obtained to enable the specifications to be met. The fourth step then, is as follows:

Step 4. The third order zero position error ITAE standard form is

$$K(s)=\frac{\omega_0^3}{s^3+1.75\omega_0 s^2+2.15\omega_0^2 s+\omega_0^3}$$

Since this standard form has three excess poles over zeros in the finite s-plane, an arbitrary term in s, say c_1s may be added to the numerator without violating the design constraint that $K(s)$ have at least two excess poles over zeros. Then, let

$$K(s)=\frac{c_1s+\omega_0^3}{s^3+1.75\omega_0 s^2+2.15\omega_0^2 s+\omega_0^3}$$

and

$$1-K(s)=\frac{s(s^2+1.75\omega_0 s+2.15\omega_0^2-c_1)}{s^3+1.75\omega_0 s^2+2.15\omega_0^2 s+\omega_0^3}$$

It should be pointed out here that $K(s)$ will yield optimum response only for $c_1=0$. However, it is reasonable to assume that relatively small nonzero values of c_1 will not significantly alter the character of the system response. Further, since c_1 is arbitrary within certain limits, it may be possible to adjust it to cause K_v to exceed the required minimum of 50/sec. It is noted that

a. $K(s)$ is stable, since it has the characteristic equation of a standard form and $1-K(s)$ has a zero at $s=0$;

b. $K(s)$ has two excess poles over zeros;

c. $K_p=\infty$ since $c_0=d_0=\omega_0^3$;

d. $K_v=\dfrac{d_0^2}{c_0d_1-c_1d_0}=\dfrac{\omega_0^6}{2.15\omega_0^5-c_1\omega_0^3}$

$$=\frac{\omega_0^3}{2.15\omega_0^2-c_1}$$

Let $\omega_0=49$ rad/sec, which is less than the specified maximum of 50 rad/sec, and let $K_v=51$/sec, which is greater than the specified minimum of 50/sec. Then, solving for c_1

$$c_1=\frac{2.15\omega_0^2K_v-\omega_0^3}{K_v}=2{,}800$$

and therefore

$$K(s)=\frac{2{,}800s+117{,}650}{s^3+85.8s^2+5{,}160s+117{,}650}$$

meets all design requirements.

The response of this system to a unit step input is shown in Fig. 4 together with the response of the ITAE third-order zero position error standard form, for which

$$K(s)=\frac{117{,}650}{s^3+85.8s^2+5160s+117{,}650}$$

and the ITAE third-order zero velocity error standard form, for which

$$K(s)=\frac{7800s+117{,}650}{s^3+85.8s^2+7800s+117{,}650}$$

As expected, the system response is different from both the standard form responses, although not significantly. For most applications the design obtained for $K(s)$ would be completely acceptable.

Step 5. By equation 1

$$D(s)=\frac{1}{G(s)}\frac{K(s)}{1-K(s)}$$

$$=\frac{s(s+a)}{K_2}\frac{2{,}800s+117{,}650}{s(s^2+85.8s+2{,}360)}$$

$$=\frac{2{,}800}{K_2}\frac{(s+a)(s+42.0)}{s^2+85.8s+2{,}360}$$

$$=\frac{2{,}800}{K_2}\frac{s^2+(42.0+a)s+42.0a}{s^2+85.8s+2{,}360}$$

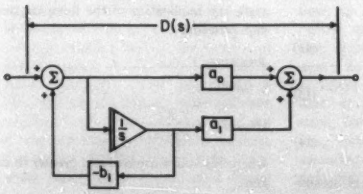


Fig. 3. Analog computer realization of D(s) for example 1

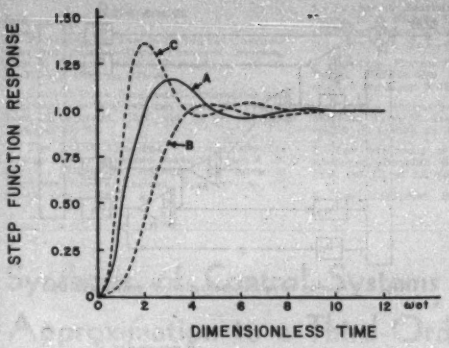


Fig. 4 (left). Step function response of: A—System of example 2, B—Optimum ITAE zero position error system, and C—Optimum ITAE zero velocity error system

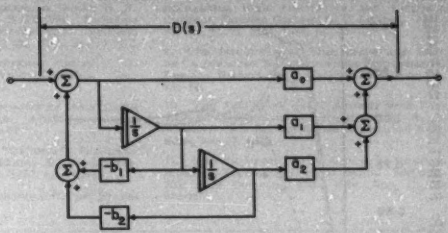


Fig. 5 (above). Analog computer realization of D(s) for example 2

Let $2{,}800/K_2=a_0$, $(42.0+a)/a_0=a_1$, $42.0-a/a_0=a_2$, $85.8=b_1$, and $2{,}360=b_2$. Then

$$D(s)=\frac{a_0s^2+a_1s+a_2}{s^2+b_1s+b_2}$$

The analog computer realization for $D(s)$ is shown in Fig. 5.

EXAMPLE 3

Given

plant $G(s)=\dfrac{1}{s(s-1)}$

Specifications: The system response to a unit step input must have the following properties:

1. Time to first zero of system error should be less than 0.2 sec.
2. Peak overshoot should be less than 50%.
3. Steady-state error should be zero.

Solution

Step 1. Stability will be assured by the use of a standard form for $K(s)$, provided that such use does not result in cancellation of singularities of $G(s)$ which lie in or very near the right half of the s-plane.

Step 2. $G(s)$ has two excess poles over zeros in the finite s-plane. Therefore $K(s)$ must have at least two excess poles over zeros.

Step 3. $G(s)$ has a pole in the right half of the s-plane at $s=1$, and a pole on the $j\omega$-axis at $s=0$. These poles must be zeros of $1-K(s)$.

Step 4. The lowest order ITAE standard form which has two excess poles over zeros in the finite s-plane is the second-order zero position error form

$$K(s)=\frac{\omega_0^2}{s^2+1.4\omega_0s+\omega_0^2}$$

and

$$1-K(s)=\frac{s(s+1.4\omega_0)}{s^2+1.4\omega_0s+\omega_0^2}$$

For $1-K(s)$ to have zeros at $s=1$ and $s=0$, it is necessary that $1=-1.4\omega_0$. This condition clearly cannot be met. The next step, as in example 2, is to try the third-order zero position error ITAE standard form with an arbitrary term in s, say c_1s, added to the numerator. Let

$$K(s)=\frac{c_1s+\omega_0^3}{s^3+1.75\omega_0s^2+2.15\omega_0^2s+\omega_0^3}$$

and

$$1-K(s)=\frac{s(s^2+1.75\omega_0s+2.15\omega_0^2-c_1)}{s^3+1.75\omega_0s^2+2.15\omega_0^2s+\omega_0^3}$$

For $1-K(s)$ to have zeros at $s=0$ and $s=1$, it is necessary that $s-1$ be a factor of

$$s^2+1.75\omega_0s+2.15\omega_0^2-c_1$$

By means of long division it may be readily shown that this condition is met, if

$$c_1=2.15\omega_0^2+1.75\omega_0+1$$

However, the condition that c_1 be relatively small no longer holds, since c_1 is greater than $d_1=2.15\omega_0^2$. More nearly optimum response will be obtained in the present example by use of the third-order zero velocity error standard form for which c_1 is not zero. Then

$$K(s)=\frac{c_1s+\omega_0^3}{s^3+1.75\omega_0s^2+3.25\omega_0^2s+\omega_0^3}$$

and

$$1-K(s)=\frac{s(s^2+1.75\omega_0s+3.25\omega_0^2-c_1)}{s^3+1.75\omega_0s^2+3.25\omega_0^2s+\omega_0^3}$$

The condition on c_1 for $s-1$ to be a zero of $1-K(s)$ now becomes

$$c_1=3.25\omega_0^2+1.75\omega_0+1$$

It is clear from an inspection of this relation that for ω_0 much greater than 1, c_1 will not differ greatly from the optimum value of $3.25\omega_0^2$, and therefore it is reasonable to assume that the transient response will not be significantly different from the optimum.

An inspection of the normalized step response for the third-order zero velocity error ITAE standard form shows that, if $\omega_0=10$ rad/sec, the rise-time and peak overshoot specifications will be met. If $\omega_0=10$

$$c_1=3.25\omega_0^2+1.75\omega_0+1=344$$

and

$$K(s)=\frac{344s+1{,}000}{s^3+17.5s^2+325s+1{,}000}$$

and

$$1-K(s)=\frac{s(s^2+17.5s-18.5)}{s^3+17.5s^2+325s+1{,}000}$$

$$=\frac{s(s-1)(s+18.5)}{s^3+17.5s^2+325s+1{,}000}$$

a. $K(s)$ is stable, since it has the characteristic equation of a standard form, and $1-K(s)$ has zero at $s=0$ and $s=1$.

b. $K(s)$ has two excess poles over zeros.

c. The specifications on time to first zero of system error and peak overshoot are met by the appropriate choice of ω_0.

d. The steady-state system error to a unit step input is zero, since $c_0=d_0=\omega_0^3=1{,}000$. Therefore all specifications are met by the system transmission function

$$K(s)=\frac{344s+1{,}000}{s^3+17.5s^2+325s+1{,}000}$$

Fig. 6 shows the unit step response for this system, together with the response of the ITAE third-order zero position and velocity error standard forms. As expected, the system response is very close to the latter.

Step 5. The compensating system $D(s)$ is readily shown to be

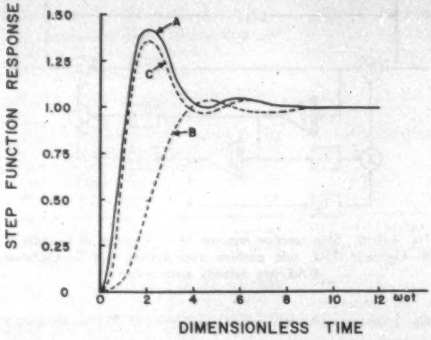


Fig. 6. Step function response of: A—System of example 3, B—Optimum ITAE zero position error system, and C—Optimum ITAE zero velocity error system

$$D(s)=\frac{1}{G(s)}\frac{K(s)}{1-K(s)}=s(s-1)\times \frac{344s+1{,}000}{s(s-1)(s+18.5)}=\frac{344s+1{,}000}{s+18.5}$$

Let

$344=a_0$, $1{,}000=a_1$, and $18.5=b_1$. Then

$$D(s)=\frac{a_0s+a_1}{s+b_1}$$

The analog computer realization of $D(s)$ is the same as for example 1 and is shown in Fig. 3.

Conclusions

A procedure for the design of duplicator control systems has been presented. While this method is by no means general in its applicability, it should be pointed out that certain of the results are quite general. Specifically the restrictions on cancellation of poles and zeros of the plant by zeros and poles of the compensating system, and the explicit relations between steady-state system error or error constants, and the coefficients of the over-all system transmission function $K(s)$, hold, no matter what design procedure is used.

An analog computer is used to realize the specified compensating system $D(s)$ as a matter of convenience, and to illustrate a possible application of such computers in control. Any other realization of $D(s)$ could be used, the choice depending on cost, convenience, etc.

The use of standard forms in the specification of the over-all transmission $K(s)$ is intended to show the relative simplicity of this little used method. Any of the well-known methods of specifying $K(s)$ may equally well be used.

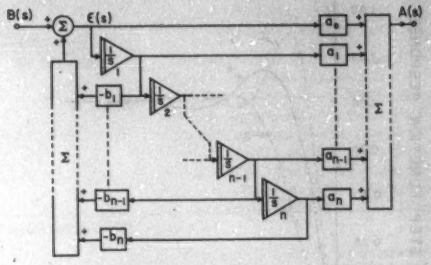


Fig. 7. Analog computer realization of generalized D(s)

Appendix I

A simple analog computer realization of a transfer function $D(s)$,[11] where

$$D(s)=\frac{A(s)}{B(s)}=\frac{a_0s^n+a_1s^{n-1}+\ldots+a_{n-1}s+a_n}{s^n+b_1s^{n-1}+\ldots+b_{n-1}s+b_n} \quad (16)$$

is shown in Fig. 7. The equations describing this system are

$$\epsilon(s)=B(s)-\epsilon(s)\sum_{i=1}^{n} b_i/s^i \quad (17)$$

$$A(s)=\epsilon(s)\sum_{i=0}^{n} a_i/s^i \quad (18)$$

Solving equation 17 for $\epsilon(s)$,

$$\epsilon(s)=\frac{B(s)}{1+\sum_{i=1}^{n} b_i/s^i} \quad (19)$$

Substituting equation 19 in equation 18

$$A(s)=\frac{B(s)\sum_{i=0}^{n} a_i/s^i}{1+\sum_{i=1}^{n} b_i/s^i} \quad (20)$$

From equation 20

$$D(s)=\frac{A(s)}{(Bs)}=\frac{\sum_{i=0}^{n} a_i/s_i}{1+\sum_{i=1}^{n} b_i/s^i}$$

$$=\frac{\sum_{i=0}^{n} a_is^{n-i}}{s^n+\sum_{i=1}^{n} b_is^{n-i}}$$

$$=\frac{a_0s^n+a_1s^{n-1}+\ldots+a_{n-1}s+a_n}{s^n+b_1s^{n-1}+\ldots+b_{n-1}s+b_n} \quad (21)$$

Appendix II

The coefficients f_i, of the Maclaurin series expansion of a function, $F(s)$

$$F(s)=f_0+f_1s+f_2\frac{s^2}{2!}\ldots \quad (22)$$

are given by

$$f_i=\left[\frac{d^iF(s)}{ds^i}\right]_{s=0} \quad (23)$$

Let $F(s)=E(s)/R(s)=1-K(s)$, where

$$K(s)=\frac{U(s)}{V(s)}=\frac{c_0s^N+c_{N-1}s^{N-1}+\ldots+c_2s^2+c_1s+c_0}{d_0s^N+d_{N-1}s^{N-1}+\ldots+d_2s^2+d_1s+d_0}$$

Then, by a straightforward application of calculus

$$f_0=[F(s)]_{s=0}=[1-K(s)]_{s=0}=1-\frac{c_0}{d_0} \quad (24)$$

$$f_1=\left[\frac{dF(s)}{ds}\right]_{s=0}=-\left[\frac{dK(s)}{ds}\right]_{s=0}=\frac{c_0d_1-c_1d_0}{d_0^2} \quad (25)$$

$$f_2=\left[\frac{d^2F(s)}{ds^2}\right]_{s=0}=-\left[\frac{d^2K(s)}{ds^2}\right]_{s=0}=\frac{2d_0(d_2c_0-d_0c_2)-2d_1(d_1c_0-d_0c_1)}{d_0^3} \quad (26)$$

The steady-state system error constants are defined by

$$\frac{E(s)}{R(s)}=1-K(s)=\frac{1}{1-K_p}+\frac{1}{K_v}s+\frac{1}{K_a}s^2+\ldots \quad (27)$$

Equating the coefficients of like powers of s in equations 22 and 27, and substituting from equations 24, 25, and 26 yields

$$K_p=\frac{1-f_0}{f_0}=\frac{c_0}{d_0-c_0} \quad (28)$$

$$K_v=\frac{1}{f_1}=\frac{d_0^2}{c_0d_1-d_0c_1} \quad (29)$$

$$K_a=\frac{2}{f_2}=\frac{d_0^3}{d_0(d_2c_0-d_0c_2)-d_1(d_1c_0-d_0c_1)} \quad (30)$$

References

1. Factors in the Design of Digital Controllers for Sampled-Data Feedback Systems, John E. Bertram. *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 151–59.

2. Sampled-Data Processing Techniques for Feedback Control Systems, Arthur R. Bergen, John R. Ragazzini. *Ibid.*, vol. 73, pt. II, Nov. 1954, pp. 236–47.

3. The Pulse Transfer Function and Its Application to Sampling Servo Systems, R. H. Barker. *Proceedings*, Institution of Electrical Engineers, London, England, *Monograph* no. 43, pt. IV, July 15, 1952.

4. Theory of Servomechanisms (book), H. N. James, N. B. Nichols, R. S. Phillips. Radiation Laboratory Series, McGraw-Hill Book Company, Inc., New York, N. Y., vol. 25, 1947, pp. 147–51.

5. Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., chaps. 5 and 1, sec. 8, 1955.

6. Control System Synthesis by Root Locus Method, Walter R. Evans. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 66–69.

7. The Synthesis of "Optimum" Transient Response, Dunstan Graham, R. C. Lathrop. *Ibid.*, vol. 72, pt. II, Nov. 1953, pp. 273–88.

8. The Transient Performance of Servomechanisms with Derivative and Integral Control, Dunstan Graham, R. C. Lathrop. *Ibid.*, vol. 73, pt. II, Mar. 1954, pp. 10–17.

9. The Influence of Time Scale and Gain on Criteria for Servomechanism Performance, Dunstan Graham, R. C. Lathrop. *Ibid.*, July, pp. 153–58.

10. The Theory of Servo Systems with Particular Reference to Stabilisation, A. L. Whiteley. *Journal*, Institution of Electrical Engineers, vol. 93, 1946.

11. Random Processes in Automatic Control (book), J. H. Laning, Jr., R. H. Battin. McGraw-Hill Book Company, Inc., 1956, App. C, pp. 376–90.

Synthesis of Control Systems Based on Approximation to a Third-Order System

C. R. HAUSENBAUER
NONMEMBER AIEE

G. V. LAGO
MEMBER AIEE

IN RECENT YEARS the high-speed performance and accuracy demanded of military applications and industrial processes have placed increasing emphasis on the dynamic characteristics of control systems. However, system design is most effective when both steady-state and transient performance specifications can be realized. Ordinarily the designer works from the open-loop to the closed-loop specifications, i.e., trial-and-error modifications are made to the fixed portion of the system until a compensation network is found that eventually leads to an acceptable closed-loop performance. In contrast, another approach, synthesis, proceeds from the closed-loop specifications to an appropriate open-loop system, which upon closure of the feedback path, satisfies the specifications. The methods of Guillemin[1] and Aaron[2] accomplish this transition, but not without difficulties and complications which arise from the fact that a closed-loop transfer function must be selected that simultaneously: 1. exhibits a pole-zero excess equal to or greater than the excess of the fixed portion of the system, and 2. satisfies a set of closed-loop specifications. Recently, Aseltine[3] introduced an inverse-root locus method that graphically determines the open-loop pole locations from a given closed-loop pole-zero configuration. However, none of these methods provides a simple procedure indicating how a closed-loop transfer function can be formulated directly from a detailed set of transient and steady-state specifications. This paper presents a quick and accurate method of synthesizing a linear, continuous, unity-feedback system from a set of specifications that eliminates the foregoing disadvantages. Attention is confined to a type I system that is compensated by an R-C realizable transfer function. To be R-C realizable, the poles of this function must be simple on the negative real axis, the origin and infinity excluded, i.e., its pole-zero excess must be equal to or greater than zero.

Theory

Terminology pertinent to the proposed method of synthesis is indicated in Fig. 1; features of the method are as follows:

1. A progression from a set of closed-loop performance specifications to an appropriate open-loop transfer function.

2. A rapid and accurate formulation from a set of specifications of a closed-loop transfer function $C(s)/R(s)$ not dependent upon the pole-zero excess of the fixed portion of the system $G_1(s)$.

3. Elimination of graphical procedures to determine the poles of the open-loop transfer function $G(s)$.

4. A simple criterion which assures that the nonzero poles of $G(s)$ are always on the negative real axis.

5. A criterion for placing a pole of $G(s)$ at a particular frequency to obtain a simplified compensation transfer function $G_1(s)$.

6. Incorporation of the $G_2(s)$ pole-zero excess requirement in a modified compensation transfer function $G_1'(s)$ as the final step of the procedure.

7. In most cases an elimination of steady-state and transient response verification of the closed-loop specifications.

Of these seven features, the first five are realizable with the simple idealized system in Fig. 1(A), idealized in the sense that $G_1(s)$ is not required to have a pole-zero excess equal to or greater than zero. Features 6 and 7 are realized with the modified system in Fig. 1(B), in which the modified compensation transfer function $G_1'(s)$ has the necessary pole-zero excess equal to or greater than zero for it to be R-C realizable.

Fig. 1(C) indicates the significance of this division of the features into distinct parts. The Bode plot for $G(s)$ represents the steady-state frequency response characteristic of an open-loop transfer function, which upon closure of the feedback path meets the steady-state and transient performance specifications, regardless of

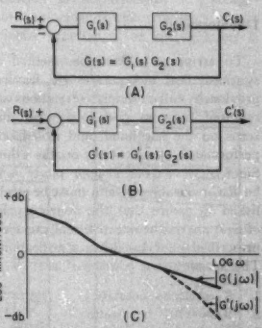


Fig. 1. System block diagrams and Bode plots illustrating proposed method of system synthesis

A—Idealized system
B—Modified system
C—Bode plots

Paper 58-797, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 29–27, 1958. Manuscript submitted October 7, 1957; made available for printing April 11, 1958.

C. R. Hausenbauer is with the University of Arizona, Tucson, Ariz., and G. V. Lago is with the University of Missouri, Columbia, Mo.

This research was supported by National Science Foundation Grant *NSF-G2304*.

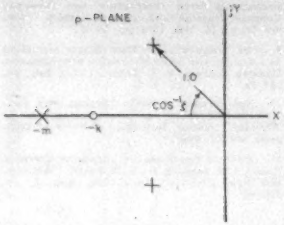


Fig. 2. Pole-zero configuration for equation 2

the nature and complexity of the fixed portion of the system. Ordinarily a relatively simple form for $G(s)$ will meet a wide variety of closed-loop specifications. The curve for the modified open-loop transfer function $G'(s)$ is identical with the one for $G(s)$ from ω equal to zero up to some frequency above the crossover frequency. The deviation from the latter curve at high frequencies represents the action of the poles that must be added to $G_1(s)$, as a consequence of the pole-zero excess of $G_2(s)$, for $G_1(s)$ to be *R-C* realizable. Use is made of the fact that the higher the frequency at which the deviation of the curves occurs, the smaller will be the effect of the added poles on the closed-loop performance of the system. As a result, consideration of the complexity of $G_2(s)$ can be delayed until the final step of the synthesis procedure, and except in a marginal situation, there is no need to determine the roots of the modified characteristic equation.

Development

To arrive at a workable method of synthesis that incorporates the features previously noted, simple correlations are required between open-loop poles and zeros on the one hand, and closed-loop performance specifications on the other. Since $G_1(s)$, or $G_1'(s)$ as appropriate, is to be *R-C* realizable, criteria must be established to assure that the nonzero poles of $G(s)$ are on the negative real axis for a prescribed set of closed-loop specifications. The specifications considered are:

1. K_s = velocity constant.
2. BW = 0.707 bandwidth.
3. γ = overshoot to unit step input, %.
4. t_r = 10% to 90% rise time.
5. t_d = 50% delay time.

To provide a means of expressing these specifications in a form suitable for comparing systems having a complex conjugate pair of poles with an arbitrary undamped angular frequency of ω_n rad/sec (radians per second), it is convenient to express transfer functions in terms of the normalized complex frequency $p = s/\omega_n = x + jy$. In effect, this transformation contracts all s-plane pole-zero positions by the factor $1/\omega_n$. Accordingly, the coordinates of the p-plane are nondimensional, and to the variable p there corresponds normalized time $t' = \omega_n t$, also nondimensional. The closed-loop specifications in normalized terminology become: 1. K_s/ω_n, 2. BW/ω_n, 3. γ%, 4. $t_r' = \omega_n t_r$, and 5. $t_d' = \omega_n t_d$. In many situations a high value of normalized velocity constant K_s/ω_n is desirable to assure that the steady-state error for a velocity input is small. Noise considerations and mechanical resonance of dynamical elements external to the feedback path require that an upper limit be placed on system bandwidth. This limit, together with the desirability of a high K_s/ω_n, suggested forming the K_s/BW ratio and incorporating it as a design criterion in the synthesis procedure.

Reference 4 presents detailed analyses of several forms of closed-loop transfer functions. The prototype of these systems is a simple second-order system that is successively made more complicated by the inclusion of a real zero, a real pole only, and lastly both a real pole and zero. Of the four systems the last one includes all the advantages and removes the disadvantages of the first three systems, both in regard to incorporating the features enumerated in the previous section and realizing an extensive variety of transient and steady-state specifications. Accordingly, the only closed-loop transfer function to be considered here is the one given by

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2 p_2}{z_1} \frac{s+z_1}{(s+p_2)(s^2+2\zeta\omega_n s+\omega_n^2)} \tag{1}$$

where p_2, z_1, and ω_n can be altered as required with the aid of the design curves to follow. In terms of the normalized complex frequency $p = s/\omega_n$, equation 1 becomes

$$\frac{C(p)}{R(p)} = \frac{m}{k} \frac{p+k}{(p+m)(p^2+2\zeta p+1)} \tag{2}$$

where $k = z_1/\omega_n$, $m = p_2/\omega_n$, and ζ is confined to the interval $0 \leq \zeta \leq 1.0$; no restriction is placed on the sign of k, but m is only positive to assure that equation 2 represents an absolutely stable system. The pole-zero configuration of this function is given in Fig. 2. The open-loop transfer function $G(p)$ is given by

$$G(p) = \frac{C(p)/R(p)}{1-C(p)/R(p)} \tag{3}$$

which upon substitution of equation 2 yields

$$G(p) = \frac{m}{k} \frac{p+k}{p[p^2+(m+2\zeta)p+(1+2\zeta m-m/k)]}$$

$$= \frac{m}{k} \frac{p+k}{p(p+p_3)(p+p_4)} \tag{4}$$

The nonzero poles of $G(p)$ are real when the inequality

$$k \leq \frac{m}{1-(\zeta-m/2)^2} \tag{5}$$

is satisfied; here k may be negative. The normalized velocity constant K_s/ω_n is obtained by substituting equation 4 in the definition

$$\frac{K_s}{\omega_n} = \lim_{p \to 0} pG(p) \tag{6}$$

with the result that

$$\frac{K_s}{\omega_n} = \frac{1}{2\zeta - \frac{1}{m} - \frac{1}{k}} = \frac{m}{p_3 p_4} \tag{7}$$

For the assumptions that the closed-loop poles are in the left-half plane and the nonzero poles of $G(p)$ are on the negative real axis, then m, p_3, and p_4 are all positive. Hence, regardless of the location of the zero of $C(p)/R(p)$, the normalized velocity constant must be a positive quantity. Clearly then, for k positive it is necessary that k satisfy the inequality

$$k \geq \frac{m}{1+2\zeta m} \tag{8}$$

to assure a positive K_s/ω_n, whereas for negative values of k, K_s/ω_n is always positive. Therefore, if inequality 5 is satisfied to assure that the nonzero poles of $G(p)$ are real, then satisfaction of inequality 8 guarantees two things: 1. the nonzero poles of $G(p)$ reside on the negative real axis, and 2. K_s/ω_n is positive. For k positive, inequalities 5 and 8 can be combined into

$$\frac{m}{1+2\zeta m} \leq k \leq \frac{m}{1-(\zeta-m/2)^2} \tag{9}$$

Fig. 3 illustrates the general nature of inequalities 5 and 9 for ζ considered as a constant. In the allowable regions A, B, C, and D, the nonzero poles of $G(p)$ are distinct and reside on the negative real axis, and K_s/ω_n is positive. However, for points exterior to or on the boundaries of these regions, the poles of $G(p)$ are complex conjugates, second order on the negative real axis, or both poles are not on the negative real axis.

To each point in an allowable region there corresponds a unique closed-loop steady-state response as well as a unique

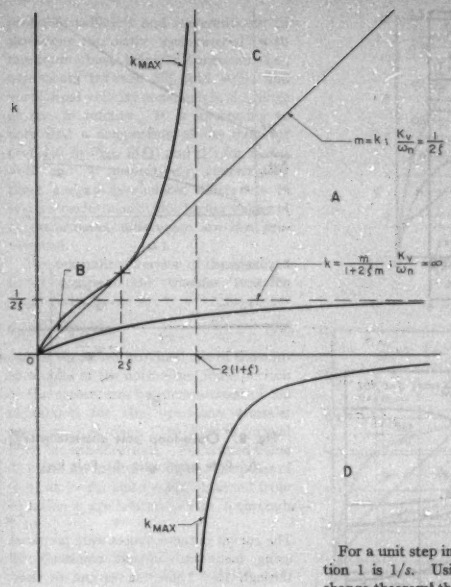


Fig. 3 (left). Regions defined by inequalities 5 and 9 in which open-loop poles are real and normalized velocity constant is positive

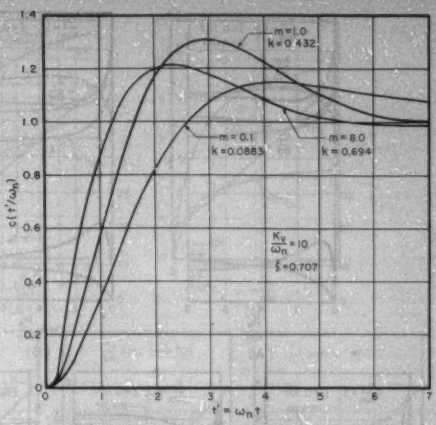


Fig. 4 (above). Closed-loop transient responses to unit step input for several values of m when $K_v/\omega_n=10$ and $\zeta=0.707$

transient response for a specified driving function. More specifically, in terms of system specifications, there exists a unique positive normalized velocity constant given by equation 7, and hence a locus along which K_v/ω_n is a constant. These loci are obtained by rearranging equation 7 as

$$k=\frac{m}{1+m\left(2\zeta-\frac{1}{K_v/\omega_n}\right)} \qquad (10)$$

in which ζ is fixed and K_v/ω_n is treated as a constant. In region A, K_v/ω_n falls in the interval $1/2\zeta \leq K_v/\omega_n \leq \infty$ for all positive values of m and k. However in the other regions, it can be shown that K_v/ω_n is always less than $1/2\zeta$, with definite restrictions on the values of m and k. Since ζ is normally confined to the interval $0.4<\zeta<1.0$, regions B, C, and D are of lesser importance. Of the four regions, region A exhibits the most favorable features and widest degree of freedom relative to a system synthesis procedure. As a consequence, the remainder of the discussion and the design curves pertinent thereto are concerned with points along loci of K_v/ω_n equal to a constant in region A only. For this region it should be noted that the zero at $(-k)$ in Fig. 2 always resides between the real pole and the origin.

For a unit step input $u(t)$, $R(s)$ in equation 1 is $1/s$. Using the familiar scale-change theorem[5] the closed-loop transient response corresponding to equation 2 is

$$c(t'/\omega_n)=1+E\exp(-mt')+F[\exp(-\zeta t')]\sin(\beta t'+\psi) \qquad (11)$$

in terms of the normalized time $t'=\omega_n t$. Expressions for E, F, β, and ψ are given in the Appendix. Fig. 4 shows several responses for movement along a line for $K_v/\omega_n=10$ when $\zeta=0.707$. For low values of m the response has a long delay time and a long time-constant tail; these features indicate the presence of an open-loop phase-lag dipole in the vicinity of the origin of the p-plane. The closed-loop normalized bandwidth BW/ω_n is obtained from

$$BW/\omega_n=v^{1/2} \qquad (12)$$

where the expression for v is given in the Appendix.

Figs. 5 through 7 set forth a graphical summary of the transient and steady-state performance characteristics of the transfer function given by equation 2.

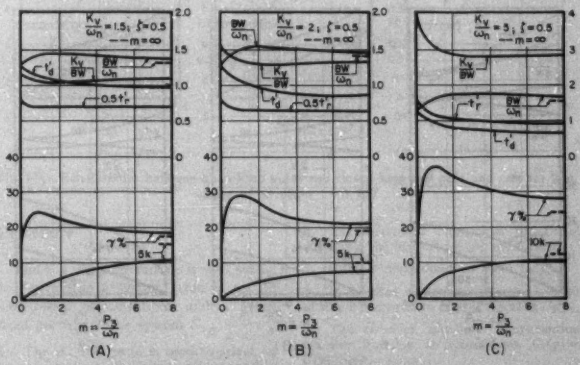


Fig. 5. Closed-loop performance characteristics when $\zeta=0.5$

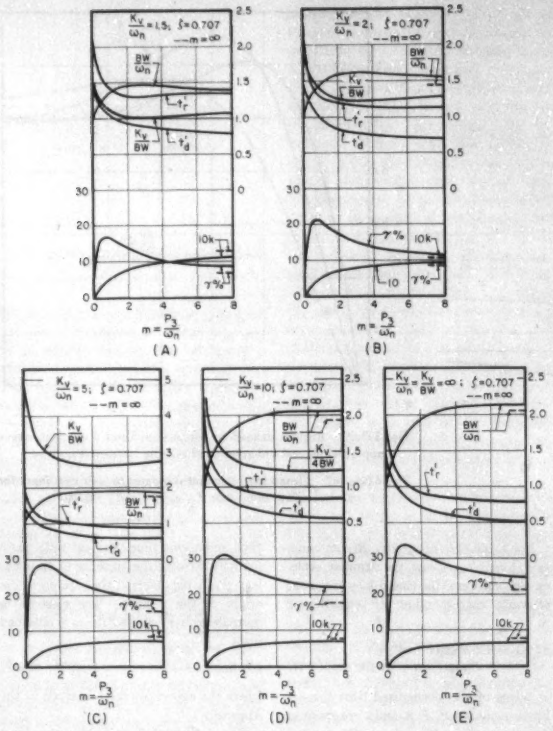


Fig. 6. Closed-loop performance characteristics when $\zeta = 0.707$

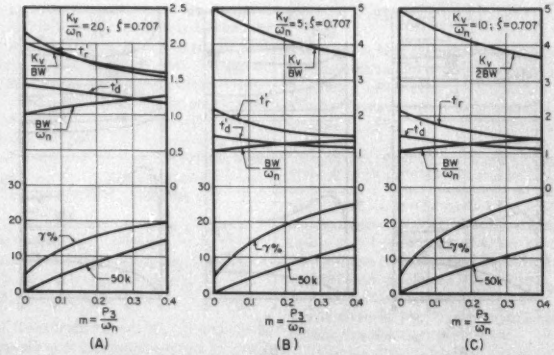


Fig. 7. Closed-loop performance characteristics for low values of m when $\zeta = 0.707$

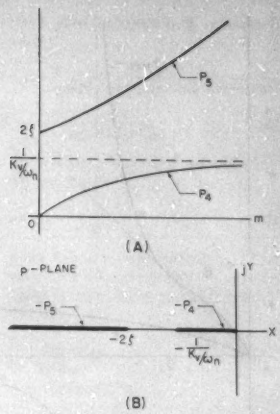


Fig. 8. Open-loop pole characteristics
A—Pole magnitudes B—Pole loci

The curves in these figures were prepared using inequality 9 and equations 10 through 12. These curves can be used to form a $C(p)/R(p)$ given by equation 2, and hence a $C(s)/R(s)$ defined by equation 1 once a value of ω_n is decided upon, from a set of closed-loop specifications. Each figure presents curves of γ, k, BW/ω_n, K_v/BW, t_r', and t_d' versus m along a line of constant K_v/ω_n when ζ is fixed at 0.5 or 0.707. Fig. 7 is concerned with values of m between zero and 0.4; as such, curves in this figure are useful for integral dipole compensation. For any point, on any curve in all figures, the nonzero poles of $G(p)$ are on the negative real axis.

In these figures, as m and k approach zero, the ratio m/k approaches unity, the pole and zero in effect cancel each other, and the system essentially reduces to a simple second-order system. To the extreme right of the figures where $m=8$, the real pole of $C(p)/R(p)$ has only a slight influence on the system performance, and the system behavior is very similar to that of a second-order system with a real zero. As m approaches infinity, the system performance is precisely that of a second-order system with a zero. Asymptotic values of γ, $10k$, and BW/ω_n are identified in the figures at $m=8.0$ by dashed lines. For intermediate values of m in the interval $0<m<8.0$, a very extensive variety of closed-loop performance is obtainable. In view of the usual tolerances on system performance specifications, the curves presented here are

probably sufficient and representative of situations normally encountered with minimum-phase transfer functions, i.e., overshoots between 5% and 40%, and normalized velocity constants in the range of one to infinity. It is interesting to note that a comparison of the data for $\zeta = 0.707$ in Figs. 6(D) and (E) for $K_v/\omega_n = 10$ and ∞ respectively, reveals that there are no appreciable differences in system performance; for higher values of ζ, performance differences are less pronounced.

The normalized version of the idealized series compensation transfer function $G_1(p)$ given by

$$G_1(p) = G(p)/G_2(p) \tag{13}$$

can be simplified provided one of the non-zero poles of the normalized fixed portion of the system can be incorporated in the expression for the open-loop transfer function $G(p)$ when $C(p)/R(p)$ is formed from the specifications. For a fixed value of ζ and an arbitrary K_v/ω_n, the poles of $G(p)$ at $(-p_4)$ and $(-p_5)$, obtained from equation 4, are related to Figs. 5 through 7 by

$$p_4 = (\zeta + m/2) - [(\zeta + m/2)^2 - (1 + 2\zeta m - m/k)]^{1/2} \tag{14}$$

and

$$p_5 = (\zeta + m/2) + [(\zeta + m/2)^2 - (1 + 2\zeta m - m/k)]^{1/2} \tag{15}$$

It is readily shown that the following inequalities apply to the situation considered here:

$$0 \leq p_4 < k < p_5 \leq \infty \tag{16}$$

and

$$0 \leq p_4 \leq \frac{1}{K_v/\omega_n} \leq 2\zeta \leq p_5 \leq \infty \tag{17}$$

Fig. 8(A) illustrates the general nature of the inequalities in expression 17 as a function of m, whereas Fig. 8(B) exhibits the same information as pole loci of $G(p)$. Consistent with satisfying a set of closed-loop specifications within prescribed tolerances, it is often possible to choose values of ζ and K_v/ω_n such that a pole of $G(p)$ can be positioned almost anywhere on the negative real axis to coincide with one of the poles of $G_2(p)$, and thereby obtain a simplified $G_1(p)$. Figs. 9 and 10, based on equations 14 and 15, present curves of p_4, p_5, and k for the particular case of $\zeta = 0.707$ along lines of constant K_v/ω_n. When used in conjunction with Figs. 6 and 7 respectively, it is often possible to quickly form a $C(p)/R(p)$, to which there corresponds an open-loop transfer function $G(p)$ containing a pole that coincides with a pole of $G_2(p)$.

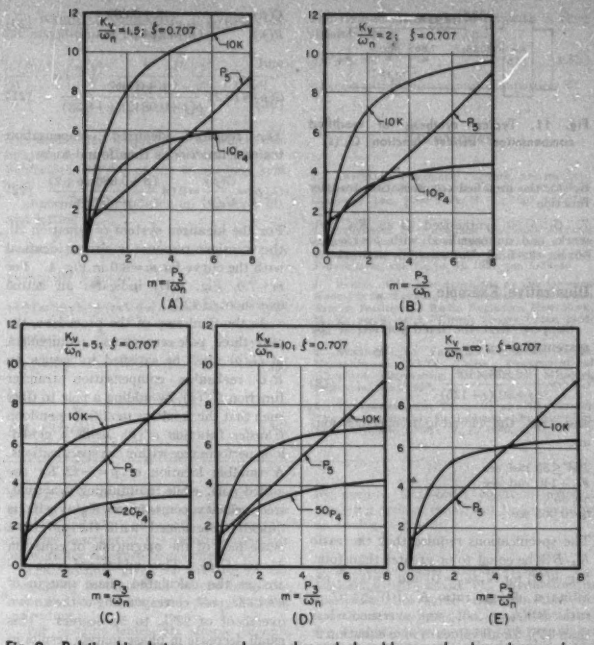


Fig. 9. Relationships between open-loop poles and closed-loop real pole and zero when $\zeta = 0.707$

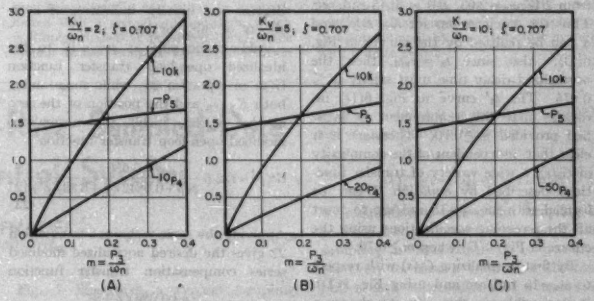


Fig. 10. Relationships between open-loop poles and closed-loop real pole and zero for low values of m when $\zeta = 0.707$

Procedure

The following procedure is well suited to synthesize a system from a set of closed-loop specifications and a given fixed portion of the system $G_2(s)$:

1. The K_v/BW ratio is used to select an appropriate set of design curves and a suitable value of ω_n for the purpose of normalization.
2. The analytical expressions for $C(p)/R(p)$ and the idealized $G(p)$ are formed.
3. $G_2(s)$ is normalized to yield $G_2(p)$.
4. If possible within specifications, $G(p)$ of step 2 is altered to exhibit a pole of $G_2(p)$.
5. The required idealized compensation transfer function is found from $G_1(p) = G(p)/G_2(p)$.
6. If necessary, poles are added to $G_1(p)$ using a Bode plot as a guide; this yields

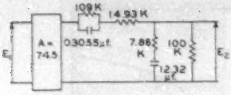


Fig. 11. Typical synthesis of modified compensation transfer function $G_1'(s)$

$G_1'(p)$, the modified compensation transfer function.

7. $G_1'(p)$ is synthesized as an R-C network, and unnormalized with $p=s/\omega_n$ to obtain the final compensation network.

Illustrative Example

Suppose that the fixed portion of the system is given by

$$G_2(s)=\frac{16.875\times10^3}{s(s+30)(s+125)} \tag{18}$$

and that the closed-loop specifications are

$BW\leq35$ rad/sec
$K_v\geq140$ rad/sec
$\gamma<35\%$
$t_d<0.065$ sec

The specifications require that the ratio K_v/BW be equal to or greater than four. Fig. 6(D) for $K_v/\omega_n=10$ and $\zeta=0.707$ exhibits a design ratio $K_v/BW\geq4.92$, a ratio $BW/\omega_n\leq2.03$, and overshoots less than 35% for all values of m in equation 2. Selecting a design value of $K_v=150$ rad/sec, then from $K_v/\omega_n=10$, $\omega_n=15$ rad/sec; from $BW/\omega_n=2.03$, $BW=30.45$ rad/sec. Thus the specifications for K_v, BW, and γ can be realized by the curves in Fig. 6(D). Also since $t_d'=\omega_n t_d$, then the normalized delay time must satisfy $t_d'\leq0.975$. The t_d' curve in Fig. 6(D) indicates that this requirement is satisfied provided $m\geq0.70$. Therefore it is clear that, independent of the complexity of $G_2(s)$, a wide variety of transfer functions described by equation 2 can be formed with an $\omega_n=15$ rad/sec to meet all the foregoing specifications using the curves in Fig. 6(D) when $m\geq0.70$.

By first normalizing $G_2(s)$ with respect to $\omega_n=15$ rad/sec and using Fig. 9(D), it is possible, in this case, to choose a pole of $G(p)$ to coincide with a pole of $G_2(p)$, and thereby obtain an immediate simplification of $G_1(p)$. Upon normalization, equation 18 becomes

$$G_2(p)=\frac{5}{p(p+2)(p+8.33)} \tag{19}$$

Equation 4 suggests selecting a value of $p_3=8.33$. For this value of p_3, the curves in Fig. 9(D) yield $p_4=0.0842$, $k=0.686$, and $m=7.0$. Accordingly equations 2 and 4 become

$$\frac{C(p)}{R(p)}=10.21\,\frac{p+0.686}{(p+7)(p^2+1.414p+1)} \tag{20}$$

and

$$G(p)=10.21\,\frac{p+0.686}{p(p+0.0842)(p+8.33)} \tag{21}$$

The required idealized compensation transfer function is then found to be

$$G_1(p)=\frac{G(p)}{G_2(p)}=2.04\,\frac{(p+0.686)(p+2)}{p+0.0842} \tag{22}$$

For the idealized system of equation 20, the transient response is almost identical with the curve for $m=8.0$ in Fig. 4. For $m=7.0$, Fig. 6(D) indicates an actual overshoot of 22%.

As the final step of the procedure, the plus three pole-zero excess requirement of $G_2(p)$ must be satisfied to assure an R-C realizable compensation transfer function $G_1'(p)$, by adding a pole to $G(p)$ such that the resulting modified open-loop transfer function $G'(p)$ yields a closed-loop performance within the specifications. A suitable location of $p=-12$ for the added pole, while maintaining a normalized velocity constant of $K_v/\omega_n=10$, is quickly determined with the aid of a Bode plot of the magnitude of equation 21 for $G(p)$. This additional pole decreases the calculated phase margin of 57.43 degrees, corresponding to the known overshoot of 22%, to 51 degrees. This small decrease in phase margin results in only a slight increase in overshoot because phase margins of more than 45 degrees are usually indicative of overshoots of 30% or less.

Adding the pole at $p=-12$ to the idealized open-loop transfer function $G(p)$ of equation 21, while maintaining both K_v/ω_n and the position of the zero at $p=-0.686$, yields the normalized modified open-loop transfer function

$$G'(p)=122.5\,\frac{p+0.686}{p(p+0.0842)(p+8.33)(p+12)} \tag{23}$$

Applying the same technique to equation 22 gives the desired normalized modified series compensation transfer function

$$G_1'(p)=24.48\,\frac{(p+0.686)(p+2)}{(p+0.0842)(p+12)} \tag{24}$$

The unnormalized version of this result is

$$G_1'(s)=24.48\,\frac{(s+10.29)(s+30)}{(s+1.263)(s+180)} \tag{25}$$

Fig. 11 gives a typical synthesis of this result. The indicated amplification is necessary because the synthesis of a transfer function by passive elements can only be realized within a gain constant. Table I lists the closed-loop specifications

Table I. Comparison of Performance Realized by Idealized and Modified Synthesis with Specifications

Closed-Loop Specifications	Idealized System	Modified System
$BW\leq35$ rad/sec	30.6	33.5
$K_v\geq140$ rad/sec	150	150
$\gamma<35\%$	22	25.8
$t_d<0.065$ sec	0.0373	0.0413
t_r* sec	0.0556	0.0546
ζ*	0.707	0.701
ω_n* rad/sec	15	16.22

* Not specified.

and makes a comparison with the closed-loop performances realized by the idealized and modified systems. The figures for the latter system are calculated values obtained by closing the feedback path on $G_1'(p)$ of equation 24. Ordinarily this need not be done; rather, an estimate of the increase in overshoot is usually satisfactory and sufficiently reliable. For this particular example, the modified closed-loop transfer function is

$$\frac{C(p)}{R(p)}=\frac{122.5(p+0.686)}{(p+5.275)((p+13.62)(p^2+1.52p+1.17)} \tag{26}$$

In terms of equation 20 and Fig. 2, the zero position is unaltered, a new pole appears at $p=-13.62$, the pole at (-7) moves to (-5.275), and the radius vectors to the conjugate poles lengthen and rotate toward the imaginary axis. As Table I indicates, the added pole tends to produce small increases in bandwidth, overshoot, and delay time accompanied by a slight decrease in rise time while maintaining a design value of velocity constant.

Conclusions

The method of system synthesis presented in this paper is based on the extensive variety of transient and steady-state performance characteristics that can be realized with a closed-loop transfer function possessing a conjugate pair of poles, a real pole, and a real zero, with the prior assurance that the poles of the corresponding open-loop transfer function are on the negative real axis. The method is direct and quick, and is sufficiently accurate for many situations normally encountered when unity feedback is applied to a minimum-phase type I system. Simplicity of the method is achieved by: 1. maintaining a design value of velocity constant, and 2. delaying consideration of the pole-zero excess of the fixed portion of the system $G_2(s)$ until the final step of the procedure. In this way

the method is essentially independent of the complexity of $G_2(s)$, and except in marginal situations, the necessity of factoring high-degree characteristic equations is avoided. In addition, a simplified *R-C* compensation network can often be realized by the expedient of pole cancellation.

Appendix

The following expressions apply to equation 11:

$$E=\frac{m-k}{k[(\zeta-m)^2+\beta^2]} \tag{27}$$

$$F=\frac{m}{k\beta}\left[\frac{(k-\zeta)^2+\beta^2}{(m-\zeta)^2+\beta^2}\right]^{1/2} \tag{28}$$

$$\beta=(1-\zeta^2)^{1/2}$$

$$\psi=\tan^{-1}\frac{\beta}{k-\zeta}-\tan^{-1}\frac{\beta}{m-\zeta}-\tan^{-1}\frac{\beta}{-\zeta} \tag{29}$$

In regard to bandwidth, the square of the magnitude of equation 2 is

$$\left|\frac{C(jy)}{R(jy)}\right|^2=\left(\frac{m}{k}\right)^2\frac{k^2+y^2}{(m^2+y^2)[(1-y^2)^2+4\zeta^2y^2]} \tag{30}$$

The normalized frequency y at which this equation has a value of 0.5 is the normalized bandwidth BW/ω_n of the closed-loop system.

Imposing this condition on equation 30, and letting

$$v=y^2=(BW/\omega_n)^2, \tag{31}$$

eventually leads to

$$v^2+(m^2+c)v^2+(m^2c+d)v-m^2=0 \tag{32}$$

where

$$c=4\zeta^2-2 \tag{33}$$

and

$$d=1-(m/k)^2 \tag{34}$$

The desired normalized bandwidth is then obtained from

$$BW/\omega_n=v^{1/2} \tag{35}$$

where v is the positive root of equation 32.

References

1. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955, pp. 278–316.

2. Synthesis of Feedback Control Systems by Means of Pole and Zero Location of the Closed Loop Function, M. R. Aaron. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1439–46.

3. Feedback System Synthesis by the Inverse Root-Locus Method, J. A. Aseltine. *Convention Record*, Institute of Radio Engineers, New York, N. Y., vol. 4, pt. 2, "Circuit Theory," Mar. 1956, pp. 13–17.

4. Synthesis of Feedback Systems, C. R. Hausenbauer. *Ph.D. Thesis*, Department of Electrical Engineering, University of Missouri, Columbia, Mo., 1957.

5. Transformation Calculus and Electrical Transients (book), S. Goldman. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1949, p. 62.

Discussion

C. J. Savant, Jr. (Northrop Aircraft, Inc., Hawthorne, Calif.): This paper presents a series of dimensionless curves which can be used to synthesize a linear and continuous control system which is approximated by a third-order equation. Although the method is limited to unity-feedback systems, it results in a synthesis procedure which displays a high degree of versatility. The method starts directly from the given closed-loop transient and steady-state performance specifications and the given fixed portion of the system. The closed-loop parameters are used directly by referring to a series of charts which yield some of the desired design quantities. The method quickly and easily produces the transfer function of the required equalizer network for the open-loop system.

The example given of the use of the method is most helpful in clarifying the use of the charts and the synthesis procedure.

For systems which are simple enough, this procedure appears to be a good approach to the synthesis problem. One of the difficulties, of course, is that a servomechanism usually can either be approximated with a second-order system, which is simpler than that presented here, or the system is too complex for approximation by even a third-order equation without incurring serious error, and hence requires an analog computer.

C. R. Hausenbauer and G. V. Lago: The authors wish to thank Dr. Savant for his discussion and agree with him that for systems too complex for an approximation by a third-order system, the use of an analog computer is highly desirable. The comments at the beginning of the second paragraph following equation 12 indicate that the curves can still be used for less complicated systems, e.g., one which can be approximated by a second-order system with a finite zero.

Analysis of Cyclic-Rate Sampled-Data Feedback-Control Systems

R. E. HUFNAGEL
NONMEMBER AIEE

IT IS NOT always possible, and sometimes not desirable, to have a sampled-data feedback-control system in which the samples are taken at uniform time intervals. Aperiodic sampled-data systems may result, for example, when time-shared computers are in the system loop or when data detection or transmission processes are inherently aperiodic (e.g. certain tracking radars).

The system discussed here is one in which the sampling rate is cyclic with a period which is an integral multiple, n, of the average time between samples. Fig. 1 illustrates a typical sampling scheme for the case $n=3$. Only the case with unity feedback and error sampling is analyzed here, since the extension of the methods described to more complicated cases is straightforward.

Development of Method

Consider the system shown in Fig. 2. Let $i(t)$ be a train of sampling pulses which multiplies the error signal, $e(t)$, to form the sampled error signal, $e^*(t)$. In order to find a closed expression for the system response it is necessary to make use of an equality of the form[1]

$$i(t)\int_0^\infty g(t-x)e(x)i(x)dx = \int_0^\infty g(t-x)i(t-x)e(x)i(x)dx \tag{1}$$

This equation is true if the sampling is periodic but in general not true if the sampling is aperiodic. It is possible, however, to view a train of cyclic-rate pulses as a superposition of n trains of periodic pulses, and therefore a solution in closed form can be obtained.

Let the times $t_1, t_2, \ldots, t_n$ and the cycle

Paper 58-803, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 3, 1958; made available for printing April 18, 1958.

R. E. Hufnagel is a graduate student at Cornell University, Ithaca, N. Y.

The author wishes to thank Prof. W. E. Meserve for his help and encouragement in the preparation of this paper; also, the National Science Foundation for the award of a fellowship for the academic year 1957–58.

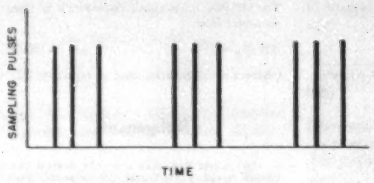


Fig. 1. Typical cyclic-rate sampling train for the case n=3

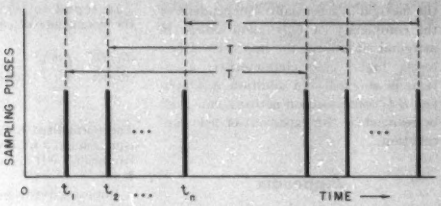


Fig. 3. Diagram showing spacing of sampling pulses

period, T, be defined as shown in Fig. 3 for a typical set of sampling impulses. For any function of time, $f(t)$, form the quantities

$$F_k(z)=\sum_{r=0}^{\infty} f(rT+t_k)z^{-r} \quad 1\leq k\leq n \tag{2}$$

Each $F_k(z)$ can often be obtained directly in closed form by using tables of modified z-transforms.[3] It is easily shown that

$$F_k(z)=z\mathfrak{F}(z,m) \qquad m=t_k/T \tag{3}$$

where $\mathfrak{F}(z,m)$ is the modified z-transform for sampling every T seconds.

Now consider the output, $c(t)$, for the system of Fig. 2 at some time $t=qT+t_j$ where q is an integer ≥ 0 and $1\leq j\leq n$. For a linear system this output, $c(qT+t_j)$, will be the sum of the effect of the open-loop transfer response, $g(x)$, operating on each of the previous error samples. That is,

$$c(qT+t_j)=\sum_{k=1}^{k=n}\sum_{r=0}^{r=q} g(rT+t_j-t_k)\times c(qT-rT+t_k) \tag{4}$$

where it has been assumed that $g(x)=0$ for $x<0$. If $g(x)$ is not a continuous function during sampling moments, it may be necessary to assume the mathematical sampling pulse to be finite in width.[3] To put this expression in convenient form let

$$G_{jk}(z)=\sum_{r=0}^{\infty} g(rT+t_j-t_k)z^{-r} \tag{5}$$

so that from equations 2, 5, and 4 one obtains

$$C_j(z)=\sum_{k=1}^{k=n} G_{jk}(z)E_k(z) \tag{6}$$

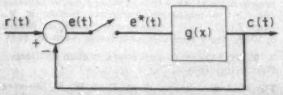


Fig. 2. Block diagram of unity feedback sampled-data system

$G_{jk}(z)$ can be expressed in closed form in terms of modified z-transforms[3] by

$$G_{jk}(z)=z\mathfrak{G}(z,m) \quad m=(t_j-t_k)/T \quad \text{for } j\geq k \tag{7A}$$

$$G_{jk}(z)=z\mathfrak{G}(z,m)-\lim_{\substack{m\to m\\ m=(t_j-t_k)/T}} z\mathfrak{G}(z,m)\times \quad \text{for } j<k \tag{7B}$$

where $\mathfrak{G}(z,m)$ is the modified z-transform of $g(x)$ with the sampling period taken as T. Equation 7(B) is necessary to insure that $g(x)=0$ for $x<0$.

For the system under consideration

$$E_k(z)=R_k(z)-C_k(z) \tag{8}$$

which combined with equation 6 gives

$$\sum_{k=1}^{k=n}[\delta_{jk}+G_{jk}(z)]C_k(z)=\sum_{i=1}^{i=n} G_{ji}(z)R_i(z) \tag{9}$$

where $\delta_{jk}=1$ if $j=k$ and $\delta_{jk}=0$ if $j\neq k$. To solve the set of equations given by equation 9 for the $C_k(z)$ it is useful to view $G_{ji}(z)$ and $\delta_{jk}+G_{jk}(z)$ as matrices operating on vectors $\mathbf{R}(z)$ and $\mathbf{C}(z)$ whose components are $R_i(z)$ and $C_k(z)$ respectively. Using matrix notation[4] equation 9 becomes

$$\|I+G(z)\|\mathbf{C}(z)=\|G(z)\|\mathbf{R}(z) \tag{10}$$

where $\|I\|$ is the unit matrix. Solving for $\mathbf{C}(z)$ gives

$$\mathbf{C}(z)=\|I+G(z)\|^{-1}\|G(z)\|\mathbf{R}(z) \tag{11}$$

Since the inverse of a matrix is equal to its adjoint matrix divided by its determinant, the condition of closed-loop stability reduces to the requirement that the zeros of the determinant $|I+G(z)|$ lie inside the unit circle. The output $c(t)$ at the times of sampling are formed from the $C_k(z)$ in the inverse manner that any $F_k(z)$ was formed from $f(t)$ in equation 2. The output between sampling instants can be found by replacing each of the $G_{ji}(z)$ in the matrix $\|G(z)\|$ of equation 11 by the appropriately modified z-transform in a manner analogous to that used in the periodic sampling case.[3]

Hold Circuits and Predictors

In equation 7 it was assumed that the open-loop response $g(x)$ was independent of the error pulse upon which it was operating. In the system shown in Fig. 4, the effect of the holding device is to clamp the signal level until the next pulse arrives. Since the duration between pulses is not uniform, the hold circuit response and hence the total open-loop response depends on the duration $t_{k+1}-t_k$. This means that in general for each value of the index k in $G_{jk}(z)$ it is necessary to use a different open-loop transfer function. For the system of Fig. 4, with a zero-order holding device, the total open-loop impulse response $g_k(x)$ associated with the kth error pulse is

$$g_k(x)=h(x)-h(x-t_{k+1}+t_k) \quad 1\leq k<n \tag{12A}$$

$$g_k(x)=h(x)-h(x-t_1+t_k-T) \quad k=n \tag{12B}$$

where

$$h(x)=\int_0^x a(y)dy \tag{13}$$

Let $H_{jk}(z)$ be the z-transform obtained by applying equation 7 to $h(x)$. The open-loop matrix elements, $G_{jk}(z)$, become

$$G_{jk}(z)=H_{jk}(z)-H_{jk+1}(z) \quad 1\leq k<n \tag{14A}$$

$$G_{jk}(z)=H_{jk}(z)-z^{-1}H_{j1}(z) \quad k=n \tag{14B}$$

These $G_{jk}(z)$ may now be used exactly as before to find the closed-loop response. In general, an analysis similar to that given here will be necessary whenever the system contains hold circuits or predictors because the outputs from these devices usually depend on the intervals, $t_{k+1}-t_k$.

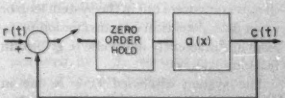


Fig. 4. Block diagram of unity feedback system with zero order hold

Illustrative Example

The problem is to determine the closed-loop response to a step input for the system of the type shown in Fig. 4 where

$$a(x)=\begin{cases}0 & x<0\\ Ke^{-x} & x\geq 0\end{cases} \qquad (15)$$

$T=1$ second $n=2$
$t_1=0$ $t_2=1/4$ second

The open-loop gain, K, will be chosen after an analysis is made of closed-loop stability. From equation 13

$$h(x)=\begin{cases}0 & x<0\\ K(1-e^{-x}) & x\geq 0\end{cases} \qquad (16)$$

The modified z-transform of $h(x)$ is[2]

$$\mathfrak{M}(z,m)=\frac{K}{z-1}-\frac{Ke^{-mT}}{z-e^{-T}} \qquad (17)$$

From equations 7 and 14

$$G_{11}(z)=K\frac{e^{-T}(e^{t_2}-1)}{z-e^{-T}}$$

$$G_{12}(z)=K\frac{1-e^{t_2-T}}{z-e^{-T}}$$

$$G_{21}(z)=Kz\frac{1-e^{-t_2}}{z-e^{-T}}$$

$$G_{22}(z)=K\frac{e^{-t_2}-e^{-T}}{z-e^{-T}} \qquad (18)$$

From equation 11

$$\begin{pmatrix}C_1(z)\\ C_2(z)\end{pmatrix}=\begin{Vmatrix}1+G_{11}(z) & G_{12}(z)\\ G_{21}(z) & 1+G_{22}(z)\end{Vmatrix}^{-1}\times \begin{Vmatrix}G_{11}(z) & G_{12}(z)\\ G_{21}(z) & G_{22}(z)\end{Vmatrix}\begin{pmatrix}R_1(z)\\ R_2(z)\end{pmatrix} \qquad (19)$$

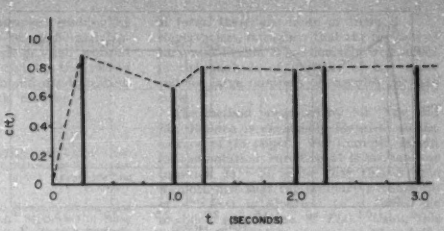
Multiplication of the second matrix by the inverse matrix gives

$$\begin{pmatrix}C_1(z)\\ C_2(z)\end{pmatrix}=\frac{1}{\Delta}\begin{Vmatrix}\Delta-1-G_{22}(z) & G_{12}(z)\\ G_{21}(z) & \Delta-1-G_{11}(z)\end{Vmatrix}\times \begin{pmatrix}R_1(z)\\ R_2(z)\end{pmatrix} \qquad (20)$$

where Δ is the determinant:

$$\Delta=[1+G(z)]=[1+G_{11}(z)]\,[1+G_{22}(z)]-G_{12}(z)G_{21}(z) \qquad (21)$$

Fig. 5. Output response to step input



The requirement for stability is that the roots of this determinant lie inside the unit circle. This expression, upon substitution of equation 18 into it, becomes the ratio of two quadratic polynomials in z and any of the standard tests[3] may be applied. The result is that for stability

$$-1<K<\frac{1-e^{-T}}{(1-e^{-t_2})(1-e^{t_2-T})} \qquad (22)$$

or substituting for T and t_2

$$-1<K<5.42 \qquad (23)$$

A reasonable K to choose to continue the example would be $K=4.0$. If the input, $r(t)$, is a unit step, equation 2 or 3 gives

$$R_1(z)=R_2(z)=z(z-1)^{-1} \qquad (24)$$

Equations 18, 20, and 24 combine to give

$$\begin{pmatrix}C_1(z)\\ C_2(z)\end{pmatrix}=\begin{Vmatrix}\dfrac{-1.450z+0.533}{z^2-0.543z+0.0644} & \dfrac{2.112z-0.777}{z^2-0.543z+0.0644}\\ \dfrac{0.884z^2-0.325z}{z^2-0.543z+0.0644} & \dfrac{-0.222z+0.079}{z^2-0.543z+0.0644}\end{Vmatrix}\times \begin{pmatrix}\dfrac{z}{z-1}\\ \dfrac{z}{z-1}\end{pmatrix} \qquad (25)$$

or

$$C_1(z)=\frac{0.662z^2-0.244z}{z^3-1.543z^2+0.607z-0.0644}$$

$$C_2(z)=\frac{0.884z^2-0.547z^2+0.079z}{z^3-1.543z^2+0.607z-0.0644} \qquad (26)$$

Expanding in powers of z^{-1} one obtains

$$C_1(z)=0.66z^{-1}+0.78z^{-2}+0.80z^{-3}+\ldots$$

$$C_2(z)=0.88+0.81z^{-1}+0.80z^{-2}+0.80z^{-3}+\ldots \qquad (27)$$

Fig. 5 shows how the two trains of pulses are combined to obtain the output.

Conclusions

A method is given for finding the closed-loop response of a cyclic-rate feed-back-control system. The actual solution of a given problem may become quite lengthy if the number of pulses in a sampling rate cycle is large, but the solution is in closed form, is exact, and has the further advantage that all the techniques of synthesis which are available with the z-transform method for periodic sampled-data systems can be applied here also.

References

1. Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955, p. 519, equation 9.51.

2. Additions to the Modified z-Transform Method, E. I. Jury. *IRE WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1957, pp. 136–56.

3. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, G. Farmanfarma. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 808–19.

4. The Mathematics of Circuit Analysis (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1949, pp. 1–71.

5. J. G. Truxal, *op. cit.*, p. 523.

Discussion

E. I. Jury and F. J. Mullin (University of California, Berkeley, Calif.): The author is to be commended for lucidly showing the application of the modified z-transform to the analysis of cyclic-rate sampled-data feedback systems. The method presented in the paper is straightforward and leads to results which are easily interpreted by those who are familiar with the z-transform and modified z-transform. In this discussion we would like to present an alternate approach to the same problem by using difference equations to describe the behavior of the system.

It is known that linear sampled-data or mixed digital-analog systems can be described by a difference equation. If the sampling rate is constant, the difference equation which describes the system has constant coefficients; if the sampling rate is periodically time varying (i.e., cyclic), the difference equation has periodically varying coefficients and the operational solution of such equations has been recently proposed by the discussers.[1]

To show the equivalence of this method to that used by the author, let us examine the illustrative example considered in the paper. The sampling pattern chosen by the author repeats itself after two samples (i.e., 1 second) and we denote the output of the system at an even number of samples by $C(2P)$ and at an odd number of samples by $C(2P+1)$, $P=0,\ 1,\ 2,\ \ldots\ \infty$. On the actual time scale $C(2P)$ corresponds to the output at $t=0,\ 1,\ 2$ seconds, ... and $C(2P+1)$ corresponds to the output at $t=0.25$ second, 1.25 seconds, 2.25 seconds.... This notation is analogous to the C_1 and C_2 used by the author. During any sampling interval, say the nth, for the $a(x)$ given by equation 15, we find the

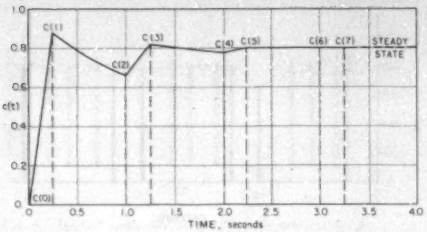


Fig. 6. The continuous response of the illustrative example

differential equation that describes the output as follows:

$$\frac{dc}{dt}+c(t)=KE(n) \tag{28}$$

$E(n)$ is the output of the zero order hold circuit shown in Fig. 4 of the paper. Solving this equation, noting that $E(n)=R(n)-C(n)$, we obtain:

$$C(t)=KR(n)[1-e^{-(t-t_n)}]+C(n)\times [e^{-(t-t_n)}-K(1-e^{-(t-t_n)})] \tag{29}$$

$$t_n\leq t\leq t_{n+1}$$

where t_n = the actual time at the start of the nth sampling interval. ($t_n=n$ seconds if n is even and $t_n=(n+0.25)$ seconds if n is odd.) If we let $t=t_{n+1}$ in equation 29 we obtain a difference equation which describes the system at the $(n+1)$th sampling instant. This equation is

$$C(n+1)=KR(n)[1-e^{-(t_{n+1}-t_n)}]+C(n)[e^{-(t_{n+1}-t_n)}-K(1-e^{-(t_{n+1}-t_n)})] \tag{30}$$

Since $t_{n+1}-t_n$ depends upon whether n is odd or even, equation 30 is a linear difference equation with periodic coefficients. It can be solved as follows. If n is even, $t_{n+1}-t_n=0.25$ second and equation 30 becomes:

$$C(2P+1)=0.222KR(2P)+C(2P)\times [0.788-0.222K] \tag{31}$$

If n is odd, $t_{n+}-t_n=0.75$ second, and equation 30 becomes:

$$C(2P+2)=0.528KR(2P+1)+C(2P+1)[0.472-0.528K] \tag{32}$$

Substituting equation 31 into equation 32 gives the following constant coefficient difference equation.

$$C(2P+2)=(0.528K)R(2P+1)+(0.105K-0.117K^2)R(2P)+(0.367-0.516K+0.117K^2)C(2P) \tag{33}$$

Using the z-transform[1] to solve equation 33, we find for $K=4.0$ and $R(2P)=R(2P+1)=1.0$

$$C(2P)=0.80[1-(0.175)^P] \tag{34}$$

We also find $C(2P+1)$, using equation 34 and equation 31, is

$$C(2P+1)=0.80[1+0.11(0.175)^P] \tag{35}$$

The foregoing two equations give the same response shown in Fig. 5 of the paper.

To obtain the response between sampling instants, we separately consider the continuous response for $P\leq t\leq(P+0.25)$ and $(P+0.25)\leq t\leq(P+1)$, $P=0, 1, 2,\ldots$

The actual continuous response for $P\leq t\leq(P+0.25)$ can be simply obtained by replacing $(t-t_n)$ in equation 29 by $mT_1=0.25m$, $0\leq m\leq 1$, and substituting equation 34 for $C(n)$. This gives

$$C(t)=4(1-e^{-0.5m})+0.80[1-(0.175)^P]\times[(e^{-0.5m}-4(1-e^{-0.25m})] \tag{36}$$

$$P\leq t\leq P+0.25$$

which, when simplified, becomes:

$$C(t)=0.80-(0.175)^P[4e^{-0.25m}-3.2] \tag{36A}$$

$$P\leq t\leq P+0.25$$

For $(P+0.25)\leq t\leq(P+1)$ we replace $t-t_n$ in equation 29 by $mT_2=0.75m$, $0\leq m\leq 1$, and substitute equation 35 for $C(n)$. This gives, after simplification:

$$C(t)=0.80+0.11(0.175)^P[4e^{-0.75m}-3.2] \tag{37}$$

$$(P+0.25)\leq t\leq(P+1)$$

The complete time response is shown in Fig. 6.

It is important to emphasize that one can always reduce difference equations with periodic coefficients to difference equations with constant coefficients[1] and the z-transform can be readily applied to the latter form as indicated in this example.

In conclusion it is hoped that this discussion will illustrate the alternate procedure which can be used to analyze this type of sampled-data feedback system. Furthermore, this method can be useful in the analysis of any process which can be described by difference equations, such as economic systems, inventory control systems, and the operational research field.

REFERENCE

1. A NOTE ON THE OPERATIONAL SOLUTION OF LINEAR DIFFERENCE EQUATIONS, E. I. Jury, F. J. Mullin. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 266, no. 3, Sept. 1958, pp. 189–205.

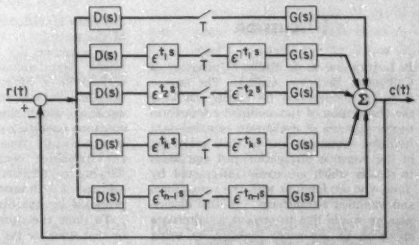


Fig. 8. Equivalent block diagram

J. Tou and E. Kinnen (Purdue University, Lafayette, Ind.): Mr. Hufnagel's paper is very interesting. The problem of cyclic-variable-rate sampled-data systems has been analyzed by an alternate method in a course on sampled-data controls taught at Purdue University. The analysis follows from techniques developed in a recent paper.[1] The method is based upon the fact that a sampler operating at a cyclic variable rate may be represented by a number of samplers operating at a constant rate, each being successively delayed by the time interval between two consecutive samples. If the cycle of the variable-rate sampler is $1/T$, the sampling period of the equivalent constant-rate samplers would be T. Assume that during the interval from 0 to T seconds the variable-rate sampler samples at the instants $t=0, t_1, t_2,\ldots t_{n-1}$. Then the sampled-data system with a cyclic-variable-rate sampler shown in Fig. 7 may be redrawn as shown in Fig. 8 with n constant-rate samplers, each being preceded by an appropriate fictitious advance element and followed by an appropriate fictitious delay element. In so doing, Fig. 8 is expressed in the basic form shown in Fig. 2 of the recent paper and in a forthcoming book.[1,2] By comparison, it is seen that

$$D_k(s)=G_c(s) \text{ and } D_k(s)=G_c(s)e^{t_k s} \tag{38}$$

$$G_k(s)=G_d(s) \text{ and } G_k(s)=G_d(s)e^{-t_k s} \tag{39}$$

Consequently, with the equivalent representation of Fig. 8, a cyclic-variable-rate sampled-data system is reduced to an equivalent constant-rate sampled-data system. Equations 20 and 21 of the reference paper

$$C(z)=\sum_{k=0}^{n-1}\frac{G_k(z)Q_k(z)}{B(z)} \tag{40}$$

$$C(z,m)=\sum_{k=0}^{n-1}\frac{G_k(z,m)Q_k(z)}{B(z)} \tag{41}$$

may then be used to determine the performance of the cyclic-variable-rate sampled-data system.

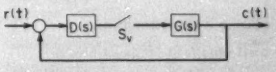


Fig. 7. Typical cyclic-variable-rate sampled-data feedback system

The example used by Mr. Hufnagel will be used to illustrate the application of equations 40 and 41. From the problem stated in equation 15 of the paper and Fig. 8, the following transfer functions can be developed:

$$D_0(s)=1,\ G_0(s)=\frac{k}{1+s}\left(\frac{1-e^{-0.35s}}{s}\right) \qquad (42)$$

and

$$D_1(s)=e^{0.35s},\ G_1(s)=\frac{k}{1+s}\left(\frac{1-e^{-0.75s}}{s}\right)e^{-0.35s} \qquad (43)$$

Equation 40 can be reduced for this example to a summation of just two terms

$$C(z)=\frac{\{RD_0(z)+RD_0(z)G_1D_1(z)-RD_1(z)G_1D_0(z)\}G_0(z)+[RD_1(z)+RD_1(z)G_0D_0(z)-RD_0(z)G_0D_1(z)]G_1(z)}{[1+G_0D_0(z)][1+G_1D_1(z)]-G_0D_1(z)G_1D_0(z)} \qquad (44)$$

By performing the required z-transformations and combinations indicated in equation 44, $C(z)$ is found equal to $C_1(z)$ in equation 26 of the paper.

$C(z,m)$ can be obtained from equation 44 by substituting $G_0(z,m)$ and $G_1(z,m)$ for $G_0(z)$ and $G_1(z)$ appearing outside the numerator brackets. The resulting $C(z,m)$ may be used to derive values of the output function at any fraction of the cyclic rate of 1 second. $C(z,\ 1/4)$ is equal to the $C_2(z)$ in equation 26 of the paper. Two other $C(z,m)$ expressions are

$$C(z,1/2)=\frac{0.795z^3-0.426z^2+0.046z}{z^3-1.54z^2+0.607z-0.0613} \qquad (45)$$

and

$$C(z,3/4)=\frac{0.719z^3-0.325z^2+0.02z}{z^3-1.54z^2+0.607z-0.0613} \qquad (46)$$

The discussion attempts to show that the concept of equivalent samplers enables the analysis and synthesis of cyclic-variable-rate sampled-data systems and related problems to be performed by the widely used z-transform and modified z-transform methods in a systematic manner.

References

1. Analysis of Sampled-Data Feedback Control Systems with Finite Sampling Duration, Julius Tou. *AIEE CP 58-801* (available on request).

2 Digital and Sampled-Data Control Systems (book), Julius T. Tou. McGraw-Hill Book Company, Inc., New York, N. Y. (to be published).

R. E. Hufnagel: I wish to thank the discussers for their comments and additions. These discussions make all the more clear that there is a great deal of interest in extending our knowledge to other than linear, time-invariant, periodically-sampled-data systems.

Mr. Jury and Mr. Mullin, in their method, take advantage of the fact that if the open-loop system corresponding to $a(x)$ is described by a differential equation of order q, the output, $C(t)$, for $t_n \leq t \leq t_{n+1}$ will depend on either the q previous samples of the output or on $C(t)$ and its first $q-1$ derivatives evaluated at the latest sampling time but on no other samples of the output. This fact, when expressed in terms of difference equations, results in z-transform equations which are simpler in form than the equations in the paper which depend on all the previous output samples. However, the number of difference equations of the type corresponding to equation 30 which must be manipulated to obtain equations corresponding to equation 33 is equal to q times the number of samples in one rate cycle whereas in my method the number of equations which must be manipulated is equal just to the number of samples in one rate cycle. Thus, although the discussers' equations are simpler in form, there are more of them if $q>1$. Nevertheless, it is clear that the method of Jury and Mullin is particularly well suited to the analysis of systems of low order and for large numbers of samples per rate cycle.

The method presented by Mr. Tou and Mr. Kinnen is essentially identical to the method of the paper. For example, except for the notation, equation 41 is the same as equation 11. To show that this is true we note first that the ordinary z-transform of $e^{mxT}F(z)$ is identical to z times the modified z-transform of $F(z)$. Using this fact, we find that for equations 42 and 43

$$RD_0(z)=R_0(z) \qquad RD_1(z)=R_0(z)$$

$$G_0(z)=G_0D_0(z)=G_{11}(z) \qquad G_0D_1(z)=G_{21}(z)$$

$$G_1(z)=G_1D_0(z)=G_{12}(z) \qquad G_1D_1(z)=G_{22}(z)$$

These expressions can be substituted into equation 44 which can then be simplified to yield the top half of equation 20 (which, of course, also contains just two terms). By setting $m=1/4$ into equation 41, one can, in a similar manner, obtain the bottom half of equation 20. Looking further, one can see that the mathematical operations necessary to analyze a system by the discussers' method are in reality the same operations necessary with the method in the paper, and that therefore the amount of labor involved in analyzing a system is essentially the same with either method. The basic difference between the method presented by Tou and Kinnen and that presented in the paper is that the discussers have manipulated a block diagram using fictitious delay and advance units to arrive at an expression for the system response while I arrived directly at an expression (equation 4) which was a clear-cut physical interpretation and which can be written by inspection.

The relative merits of each of the methods presented here will best be revealed with time and experience.

Multivariable Control System Synthesis

R. J. KAVANAGH
ASSOCIATE MEMBER AIEE

Synopsis: An important problem in the field of multivariable control systems is the design of suitable feedback and series compensation elements which, when added to a given system, result in an over-all system with improved characteristics. This improvement may apply to the behavior of the controlled variables, to the disturbance response of the system, or to both simultaneously.

This paper presents some methods which may be used to obtain the transfer matrices of the feedback and series elements under these circumstances. Restrictions on the transfer matrices of the desired over-all systems are developed and tabulated together with design equations for three synthesis methods.

THE type of system studied is illustrated in Fig. 1. In this figure, H is the given system which is assumed to be linear, and Q is a device whose output signals z are linear combinations of the x and y signals. This form of modification involves, in general, both feedback and operation on input sigals. It is to be noted that no operations are performed on the given system's outputs. This is in accordance with the fact that, in practice, the outputs of control systems are not in a form which may be operated upon. This restriction however does not prevent the measurement of these output variables which is necessary for the use of feedback.

By use of the Laplace transformation, the system of Fig. 1 may be described by the following equations in which the variables and coefficients are functions of the complex frequency variable:

$$\begin{aligned} y_1 &= h_{11}z_1+h_{12}z_2+\ldots+h_{1n}z_n \\ y_2 &= h_{21}z_1+h_{22}z_2+\ldots+h_{2n}z_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m &= h_{m1}z_1+h_{m2}z_2+\ldots+h_{mn}z_n \end{aligned} \qquad (1)$$

Paper 58-852, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 7, 1958; made available for printing April 28, 1958.

R. J. Kavanagh is with the University of Toronto, Toronto, Ont., Canada.

The author wishes to acknowledge the guidance received from Prof. J. M. Ham of the University of Toronto.

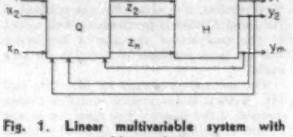


Fig. 1. Linear multivariable system with modifying device

$$z_1 = c_{11}x_1 + c_{12}x_2 + \ldots + c_{1n}x_n + k_{11}y_1 + k_{12}y_2 + \ldots + k_{1m}y_m$$

$$z_2 = c_{21}x_1 + c_{22}x_2 + \ldots + c_{2n}x_n + k_{21}y_1 + k_{22}y_2 + \ldots + k_{2m}y_m$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$z_n = c_{n1}x_1 + c_{n2}x_2 + \ldots + c_{nn}x_n + k_{n1}y_1 + k_{n2}y_2 + \ldots + k_{nm}y_m \quad (2)$$

In matrix form these become:

$$Y = HZ \quad (3)$$

$$Z = CX + KY \quad (4)$$

It may be seen, therefore, that the device Q is completely described by the two transfer matrices C and K. By combining equations 3 and 4, the following is obtained:

$$Y = H'X \quad (5)$$

where

$$H' = HC + HKH' \quad (6)$$

or

$$H' = (I - HK)^{-1}HC \quad (7)$$

provided that $(I - HK)$ is nonsingular. Thus a relationship has been obtained between the given transfer matrix H and the new transfer matrix H' in terms of the transfer matrices describing Q.

The synthesis problem is therefore seen to consist of the selection of suitable matrices for C and K in order that a desired system transfer matrix H' may be obtained. It has been previously shown[1] that there are, in general, an infinite number of C and K matrices which satisfy the specification. To obtain any one of these solutions it is necessary to apply constraints to C and K. These may be applied in two ways. In some cases it may be possible to constrain certain elements of C and K such that a unique solution for the remaining elements is obtained. Such constraints may arise from the fact that certain of the variables concerned may be disturbances and hence may not be operated upon. This technique is illustrated in reference 1.

The second method of applying constraints involves the selection of certain forms of the matrices C and K rather than the constraining of certain elements. This paper describes four methods of synthesis based on the restriction of C and K to certain forms.

Closed-Loop Synthesis

As the heading of this section implies, the methods to be described involve the use of feedback. The distinction between open- and closed-loop systems may be seen to depend upon the nature of the matrix K. If K is constrained to be a null matrix, the resultant system is an open-loop system, otherwise the system is closed-loop since feedback is involved.

Before describing closed-loop synthesis methods it should be pointed out that open-loop synthesis is completely valid for some purposes. In this case, K is constrained to be a null matrix. Hence equation 6 becomes

$$H' = HC \quad (8)$$

Thus, if H is a nonsingular matrix, C is uniquely given by

$$C = H^{-1}H' \quad (9)$$

This method has been described elsewhere[2] and is not described further in this paper.

Apart from the open-loop technique of synthesis there are three convenient closed-loop techniques. These are based on the following constraints on C and K:

Method I: $K = I$ (the identity matrix)
Method II: $C = I$
Method III: $C = -K$

These constaints are permissible since, in all cases, the resultant matric design equation contains at least as many unknowns as equivalent simultaneous equations.

It is interesting to compare the systems obtained when these constraints are applied with the corresponding single-variable systems. Method I results in a system in which the fed-back output is unmodified and is added to the input to form the basic system input. The system of method II is one in which the fed-back output is modified and added to the unmodified input. Finally, the third method results in the output being subtracted from the input and then modified inside the loop. It is apparent that three additional techniques result from the constraints $K = -I$, $C = -I$, and $C = K$. These give alternative designs which may, in some cases, give more desirable systems than the methods described in this paper.

The design equations that result from the use of the first three constraints are shown in Table I. These design equations all result from the substitution of the appropriate constraint into equation 6. It may be seen that only method II is applicable to nonsquare systems. This is because, with reference to equation 2, the constraints $K = I$ and $K = -C$ imply that the system is square. Because of the nonexistence of inverses of singular matrices, the equations for singular and nonsingular systems differ. It is now shown that the use of these synthesis methods imposes some conditions on the types of systems which may be synthesized using a given H.

Restrictions on H'

It has been previously shown[2] that the use of the configuration of Fig. 1 automatically results in a restriction on the form of H'. The constraints implied in the use of the three methods of closed-loop synthesis result in additional restrictions in some cases.

In general, a system with transfer matrix H may be modified to give a system with transfer matrix H' by use of the configuration of Fig. 1 if and only if it is possible to transform the matrix H to H' by a series of column operations. The allowable column operations are:

1. Multiplication of a column by any ratio of two polynomials in the complex frequency having integral powers, or by any real number including zero.
2. Addition of any multiple, as in item 1, of a column to a second column.
3. Interchange of any two columns.

Note that the physical realizability of H' is not being considered at this point. This basic restriction is referred to as "H is column equivalent to H'." The equivalence referred to here is, in general, asymmetric.

Consider method I for a nonsingular H. Since C is uniquely determined by the design equation, it follows that there are no additional restrictions on H'. For any H', the design equation gives the desired C matrix. Hence the realizability

Table I. Design Equations

Method	Constraint	Square Nonsingular H	Square Singular H	H Not Square
I	$K = I$	$C = (H^{-1} - I)H'$	$HC = (I - H)H'$	
II	$C = I$	$K = H^{-1} - H'^{-1}$	$HKH' = H' - H$	$HKH' = H' - H$
III	$K = -C$	$C = H^{-1}H'(I - H')^{-1}$	$HC = H'(I - H')^{-1}$	

Table II. Restrictions on H'

Method	Restrictions on H'		
	Square Nonsingular H	Square Singular H	H Not Square
I	H is column equivalent to H'	H is column equivalent to H'	
II	H' must be nonsingular	H is column equivalent to H'; H' is row equivalent to H; rank H = rank H' (necessary but not sufficient)	H is column equivalent to H'; H' is row equivalent to H
III	H is column equivalent to H'; $(I-H')$ must be nonsingular	H is column equivalent to H'	rank H = rank H' (necessary but not sufficient)

criterion is that H is column equivalent to H'.

Consider method I for a singular H. The design equation may be written as

$$H' = H(C+H') \tag{10}$$

If H is singular, then the right-hand side of equation 10 is singular and hence H' must also be singular. The postmultiplication of H by the matrix $(C+H')$ is equivalent to the modification of H by a set of column operations. Hence there is no additional restriction on H' other than that it must be possible to obtain H' from H by a series of column operations. Method I, then, implies no extra restrictions on H' for either a nonsingular or a singular H.

Consider method II for a nonsingular H. Substitution of $C=I$ in equation 6 yields:

$$H' = H+HKH' \tag{11}$$

which becomes

$$H = (I-HK)H' \tag{12}$$

Now if H is nonsingular, the right-hand side of equation 12, and hence H', must be nonsingular. An additional restriction on H' in this case is therefore that it must be nonsingular. This fact is the justification for writing the design equation for this case in terms of H'^{-1}.

Consider method II for a singular H. Because of the basic restriction resulting from the use of the configuration of Fig. 1, it follows that H' must also be singular. The design equation for this case may be written as

$$(I-HK)H' = H$$

The premultiplication of H' by $(I-HK)$ is equivalent to a set of row operations on H'. Thus, in addition to the basic restriction, it follows that it must be possible to transform H' to H by a set of row operations of the same type used for column operations. Hence, in this case, H' has an additional restriction as well as the basic one. A further result may be obtained by combining these two restrictions. From the first it follows that

$$\text{rank } H \geq \text{rank } H'$$

From the second it follows that

$$\text{rank } H \leq \text{rank } H'$$

Hence, for this case it is a necessary, but not sufficient, condition that

$$\text{rank } H = \text{rank } H'$$

Consider method III for a nonsingular H. Substitution of $K=-C$ in equation 6 leads to

$$H' = HC-HCH' \tag{13}$$

This may be written as

$$H^{-1}H' = C(I-H') \tag{14}$$

The left-hand side of equation 14 is clearly nonsingular. Hence both terms, and in particular $(I-H')$, must also be nonsingular. The additional restriction in this case, then, is that $(I-H')$ must be nonsingular.

Consider method III for a singular H. The substitution of $K=-C$ in equation 6 yields

$$H' = HC(I-H') \tag{15}$$

The postmultiplication of H by $C(I-H')$ is equivalent to a set of column operations. Hence no additional restriction arises in this case. The form of the design equation given in Table I is justified since in this case H' is always singular and hence $(I-H')$ is always nonsingular.

The necessary restrictions on H' for all cases of the three closed-loop synthesis method have now been determined. A summary of these results is contained in Table II.

Example of System Design

Turbopropeller Engine Control

The operation of a turbopropeller engine in the region of a set operating point can be described by the following equations[3,4]:

$$y_1(s) = \frac{-a}{1+ts}z_1(s) + \frac{b}{1+ts}z_2(s)$$

$$y_2(s) = \frac{as}{1+ts}z_1(s) + \frac{(c-be)+cts}{1+ts}z_2(s) \tag{16}$$

where

$y_1(s)$ = Laplace transform of deviation of engine speed from its normal value

$y_2(s)$ = Laplace transform of deviation of turbine inlet temperature from its normal value

$z_1(s)$ = Laplace transform of deviation of propeller blade angle from its normal value

$z_2(s)$ = Laplace transform of deviation of fuel rate from its normal value

a, b, c, e = positive real constants

t = engine time constant

For the purposes of computation, assign arbitrary values to a, b, c, e, and t. These values may not be representative of any particular engine, but simplify the design computations. These assumptions do not in any way affect the generality of the methods to be used. Let the resultant equations become

$$y_1(s) = \frac{-2}{1+s}z_1(s) + \frac{3}{1+s}z_2(s)$$

$$y_2(s) = \frac{4}{1+s}z_1(s) + \frac{2+8s}{1+s}z_2(s) \tag{17}$$

Thus for this system

$$H = \begin{vmatrix} \frac{-2}{1+s} & \frac{3}{1+s} \\ \frac{4}{1+s} & \frac{2+8s}{1+s} \end{vmatrix} \tag{18}$$

Let it be desired to obtain a modified system whose transfer matrix is given by

$$H' = \begin{vmatrix} \frac{-2}{1+s} & 0 \\ 0 & \frac{2+8s}{1+s} \end{vmatrix} \tag{19}$$

This results in a system whereby the engine speed and temperature may be independently controlled.

Since the system is square, all of the three methods are eligible for this synthesis. To determine the ranks of H and H', substitute $s=0$ in the matrices. It may be seen that both H and H' are nonsingular. It immediately follows from Table II that both methods I and II may be used in this synthesis. Finally, to decide on the eligibility of method III, it is necessary to form the matrix $(I-H')$

$$(I-H') = \begin{vmatrix} \frac{3+s}{1+s} & 0 \\ 0 & \frac{-1-7s}{1+s} \end{vmatrix} \tag{20}$$

By substitution of $s=0$, it is found that $(I-H')$ is nonsingular. Hence method III is also eligible. Since all three methods may be used, the synthesis will be accomplished by the three different methods.

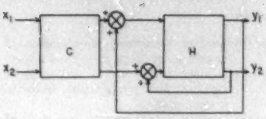


Fig. 2. Turbopropeller engine control system, method I

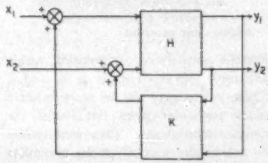


Fig. 3. Turbopropeller engine control system, method II

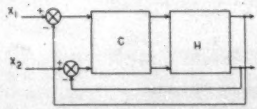


Fig. 4. Turbopropeller engine control system, method III

METHOD I

From Table I, the design equation is

$$C=(H^{-1}-I)H' \tag{21}$$

Substitution of the appropriate matrices in this equation gives

$$C=\begin{Vmatrix} \dfrac{9(1+4s/9)}{4(1+s)} & \dfrac{3(1+4s)}{8(1+s)} \\ -\dfrac{1}{2(1+s)} & -\dfrac{7(1+4s)}{4(1+s)} \end{Vmatrix} \tag{22}$$

This system is shown in block-diagram form in Fig. 2.

METHOD II

From Table I, the design equation is

$$K=H^{-1}-H'^{-1} \tag{23}$$

Substitution for H and H' yields

$$K=\begin{Vmatrix} \dfrac{3}{8} & \dfrac{3}{16} \\ \dfrac{1}{4} & -\dfrac{3}{8(1+4s)} \end{Vmatrix} \tag{24}$$

This system is shown in Fig. 3.

METHOD III

From Table I, the design equation is

$$C=H^{-1}H'(I-H')^{-1} \tag{25}$$

Substitution of the appropriate matrices yields

$$C=\begin{Vmatrix} \dfrac{1+4s}{12(1+s/3)} & -\dfrac{3(1+4s)}{8(1+7s)} \\ -\dfrac{1}{6(1+s/3)} & \dfrac{1+4s}{4(1+7s)} \end{Vmatrix} \tag{26}$$

This system is shown in Fig. 4.

Systems with Disturbance Inputs

It is frequently necessary to consider the effects of disturbance inputs as well as controlling inputs on the controlled variables in a system. In the multivariable case this is equivalent to considering the system matrix to have additional columns corresponding to the disturbances. It is desired to be able to design the over-all system to have acceptable disturbance-output response as well as acceptable input-output response. Two possible approaches to this problem are considered.

The first method assumes that the disturbance inputs are just like any other types of inputs. The basic system transfer matrix will then have a total of $(n+d)$ inputs and m outputs where n is the number of controlling inputs, d is the number of disturbance inputs, and m is the number of controlled outputs. Hence the transfer matrix will normally have more columns than rows. The design problem reduces to the determination of a suitable C and K matrix such that the over-all system transfer matrix has desirable elements.

Unfortunately, because of the rectangular nature of H in this case, only method II of those previously given is applicable. Furthermore, method II is characterized by rather limiting restrictions. These restrictions limit the application of this method and an alternative method is desirable.

The second method postulates that the basic system has a block diagram as shown in Fig. 5. In this diagram, Z represents the controlling inputs, D the disturbance inputs, and Y the controlled outputs. This is a completely general representation and any basic multivariable system may be thus represented provided H_1 and H_2 are chosen appropriately. The system may be described by the following matric equation

$$Y=H_2H_1Z+H_2D \tag{27}$$

Feedback may now be applied around this system such that the signals Z are linear combinations of the outputs Y and new over-all system inputs X. The new system is shown in Fig. 6. It follows that

$$Z=CX+KY \tag{28}$$

and

$$Y=(I-H_2H_1K)^{-1}H_2H_1CX+(I-H_2H_1K)^{-1}H_2D \tag{29}$$

providing $(I-H_2H_1K)$ is nonsingular.

Let the desired response of the system be

$$Y=AX+BD \tag{30}$$

Equating matric coefficients gives

$$A=(I-H_2H_1K)^{-1}H_2H_1C \tag{31}$$

$$B=(I-H_2H_1K)^{-1}H_2 \tag{32}$$

The solution of these equations for C and K gives the desired system. In this discussion, the case where the number of inputs (Z and D) and outputs Y are identical is treated. This restriction has the justifications of being in accordance with the normal practical case and of giving a useful result in a useful form with a simple derivation. It will, of course, be normally possible to determine C and K for other cases although the solutions will not generally be unique.

On the basis of this restriction, equation 32 yields a unique solution for K

$$K=H_1^{-1}(H_2^{-1}-B^{-1}) \tag{33}$$

providing H_1, H_2, and B are nonsingular.

Substitution of equation 33 into equation 31 yields a unique solution for C

$$C=H_1^{-1}B^{-1}A \tag{34}$$

Thus, the important conclusion is reached that the disturbance response and the input-output response of the system may be chosen independently. As a special case of this system, it also follows that independent design for disturbance response and input-output response may be achieved in single-variable systems as has been demonstrated previously by Lang and Ham,[5] and Truxal.[6]

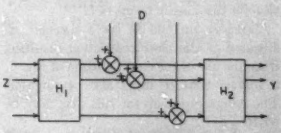


Fig. 5. Basic multivariable system with disturbance inputs

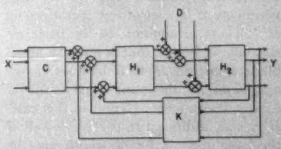


Fig. 6. Multivariable system with disturbance inputs after modification

Example of System Design

Consider a hypothetical system in which there are two controlling inputs, two outputs, and two disturbance inputs, and where

$$H_1 = \begin{Vmatrix} -\frac{2}{1+s} & \frac{3}{1+s} \\ \frac{4}{1+s} & \frac{2+8s}{1+s} \end{Vmatrix} \tag{35}$$

$$H_2 = \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix} \tag{36}$$

Let it be desired to obtain a system whose disturbance response is described by

$$B = \begin{Vmatrix} \frac{s}{1+2s} & 0 \\ 0 & \frac{s}{1+2s} \end{Vmatrix} \tag{37}$$

and whose input-output response is described by

$$A = \begin{Vmatrix} -\frac{2}{1+s} & 0 \\ 0 & \frac{2+8s}{1+s} \end{Vmatrix} \tag{38}$$

Hence the disturbance response is improved and simultaneously noninteracting controls are obtained.

For this system it follows that

$$H_1^{-1} = \begin{Vmatrix} -\frac{1+4s}{8} & \frac{3}{16} \\ \frac{1}{4} & \frac{1}{8} \end{Vmatrix} \tag{39}$$

$$H_2^{-1} = \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix} \tag{40}$$

$$B^{-1} = \begin{Vmatrix} \frac{1+2s}{s} & 0 \\ 0 & \frac{1+2s}{s} \end{Vmatrix} \tag{41}$$

From equation 33 and equation 34 the following matrices are obtained for C and K

$$C = \begin{Vmatrix} \frac{2(1+4s)(1+2s)}{8s(1+s)} & \frac{3(1+2s)(2+8s)}{16s(1+s)} \\ \frac{-(1+2s)}{2s(1+s)} & \frac{(1+2s)(2+8s)}{8s(1+s)} \end{Vmatrix} \tag{42}$$

$$K = \begin{Vmatrix} \frac{(1+4s)(1+s)}{8s} & \frac{-3(1+s)}{16s} \\ \frac{-(1+s)}{4s} & \frac{-(1+s)}{8s} \end{Vmatrix} \tag{43}$$

Conclusions

This paper has described methods by which a given linear multivariable system may be modified in a logical manner so that a system with different characteristics is obtained. The application of these ideas to the control of disturbances in multivariable systems has also been illustrated.

The methods described here will be of use in the synthesis of systems involving more than one variable where the operation of the system may be taken as linear. Under these conditions synthesis of desirable controlling input-output and disturbance input-output responses may be accomplished simultaneously.

References

1. The Application of Matrix Methods to Multi-Variable Control Systems, R. J. Kavanagh. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 262, Nov. 1956, pp. 349–67.

2. A Control System Interpretation of Certain Properties of Matrices, R. J. Kavanagh. *The Matrix and Tensor Quarterly*, Tensor Club of Great Britain, Kent, England, vol. 8, no. 2, Dec. 1957, pp. 32–38.

3. General Algebraic Method Applied to Control Analysis of Complex Engine Types, A. S. Boksenbom, R. Hood. *NACA Report 980*, National Advisory Committee for Aeronautics, Washington, D. C., 1950.

4. Engineering Cybernetics (book), H. S. Tsien. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 63–65.

5. Conditional Feedback Systems—A New Approach to Feedback Control, G. Lang, J. M. Ham. *AIEE Transactions*, vol. 74, pt. II, July 1955, pp. 152–61.

6. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., 1955, pp. 320–21.

Discussion

F. William Nesline, Jr. (Massachusetts Institute of Technology, Cambridge, Mass.): Dr. Kavanagh is to be commended for presenting a most readable account of some of the problems of multivariable control system design. The philosophy behind this paper, and all the other published papers on this topic, is that of cancellation compensation; i.e., the design of a controller to cancel the undesired dynamics of the fixed element or plant and to substitute the desired dynamics. If the desired dynamics H' are properly chosen, this concept can be applied and a satisfactory design achieved. However, if the desired dynamics are not properly chosen, the resulting design may yield a completely unsatisfactory system. Therefore, the crucial step in this type of design is the choice of the desired dynamics H'. In the case of multi-input single-output systems, considerable attention has been given to this problem by Newton[1] in his extension of Wiener's theory of the optimum linear filter.[2] The three items that must be considered in choosing H' are:

1. The form of the fixed element (plant) transfer functions.
2. The saturation characteristics of the fixed element.
3. The noise characteristics of the input signal (x).

For example, if the transfer function of the fixed element has a zero in the right half of the s-plane and the desired dynamics do not include this zero, the cancellation compensator is *not* realizable. Also, if the transfer function of the fixed element has more poles than zeros, as is always the case, and the desired dynamics do not have at least as many more poles than zeros, the controller will have more zeros than poles. Such a transfer function can be realized only over a limited range and even if it is realized, its output signals may be large enough to cause serious saturation problems in the fixed element. Finally, if noise is present at the input x to the system, the desired dynamics must represent a compromise between noise transmission and desired signal transmission and will therefore depend upon the characteristics of the desired signal and its noise, as well as the characteristics of the fixed element.

While the effects listed in the foregoing have been investigated rather extensively[1] for single-output systems, no comparable investigation has been made in the case of truly multivariable systems. Even though it is unlikely that startlingly new qualitative ideas would stem from such an investigation, the results ought to systematize the process of choosing a desired system matrix.

References

1. Analytical Design of Linear Feedback Controls (book), G. C. Newton, Jr., L. A. Gould, J. F. Kaiser. John Wiley & Sons, Inc., New York, N. Y., 1957.

2. Extrapolation, Interpolation, and Smoothing of Stationary Time Series (book), N. Wiener. *Ibid.*, 1949.

R. J. Kavanagh: Prof. Nesline's comments are most appropriate and one must certainly agree with him that the choice of a suitable H' matrix is a crucial step in the design of a multivariable control system. This paper is based on the supposition that the H' matrix has already been determined in some manner and hence it consists of what might be called the second stage of the synthesis. The first stage, the choice of H', is a separate problem whose solution will involve consideration of the items listed by Prof. Nesline.

A further item may be added to Prof. Nesline's list when one considers a given system H for which either the elements are time dependent, or the elements are known only within a limited accuracy. In such a case, H' will differ from its nominal value and the design problem will then involve the choice of a suitable controller to yield an H' which will remain within prescribed limits as well as being satisfactory with regard to over-all system performance.

The technique of selecting a suitable H' matrix is a problem which certainly deserves further investigation either by an extension of single-variable system trial-and-error methods or by the use of an analytical design procedure.

A Simple Nonlinearized Control System

J. ZABORSZKY
MEMBER AIEE

A LINEAR control system is possibly the best compromise solution for an environment which includes a more or less random assortment of all kinds of inputs. However, for control systems which work with only one or two well-defined kinds of inputs such as steps and ramps the linear system has serious limitations; mainly:

1. The maximum available torque or force of the servomotor is utilized only throughout a small fraction of the operating cycle, so the efficiency is low and the response inherently slow.

2. The transient response has a theoretically infinite duration, and practically it consists of a long trail of damped oscillations or a long asymptotic tail.

This situation can be remedied by purposefully introducing nonlinearities into the system. By properly selecting such nonlinearities, the reponse of a system can be optimized in speed, accuracy, and efficiency for a specific type of input, but there is a price to pay for this improvement in a deterioration of the response to other types of inputs. This price, however, may be well worth paying if such other types of inputs do not occur in the environment of the system.

This paper presents a simple means of nonlinearization, which accomplishes the optimization of the response to displacement steps with a minimum of equipment and circuitry. In the latter it is superior in many cases to previously proposed systems, some of which[1-6] are quite similar in nature to the system proposed here, at least as applied to second-order servomechanisms.

How Nonlinearities Can Optimize Response

During the transient response of a second-order linear instrument servomechanism, oscillations may be caused by the fact that in such systems maxima and zeros of error (or potential energy) and error rate (or kinetic energy) respectively are alternating and interlaced so that maximum error coincides with zero error rate or vice versa. Consequently, zero error and zero error rate do not coincide until infinite time when all transient energy is dissipated. Yet simultaneous zero error and error rate are the conditions for the completion of the response in a second-order system. In a third-order system, error, error rate, and error acceleration must be simultaneously zero with a displacement step input etc.

If the error and error rate can be simultaneously brought to zero after a displacement step input in a second-order (zero-displacement error-type) system within a finite time interval, then this finite interval is the duration of the response; the system remains at rest consecutively until the next input.

The aim is usually attained by nonlinearization or more precisely by producing two consecutive linear responses which jointly bring the system to an instant of simultaneous zero error and zero error rate. This requires some kind of switching in one or two steps. These switchings result in steplike changes of one or more parameters, such as damping The ultimate of these techniques[7-8] as far as speed of response is concerned is the off-on type of servomechanism which first applies full torque ahead and then full torque reverse. This result is the fastest response obtainable using a given servomotor.

The Proposed System

The common drawback of most of the previously described sectionally linear and relay or on-off types of servomechanisms is the rather complex switching and monitoring equipment required to switch at the right time with the right amount to bring the system into zero error and zero error rate at a future finite time. This complexity may be avoided simply by making use of the fact that the peak time and percentage overshoot in the response of a linear system to a step input are independent of the height of the step input and depend only on the system itself. At the displacement peak during the transient the error rate is zero, so if at this time the error is reduced to zero the second-order system will remain at rest from there on.[3,4] However, the true error at the first transient peak will be actually zero if the input during the first transient half-cycle was reduced in the proportion of the final output to the first transient peak of the step response.

If the normalized closed-loop transfer function of the system, Fig. 1, without the timer is

$$\frac{C(s)}{R_0(s)}=\frac{1}{s^2+2\zeta s+1} \tag{1}$$

then the peak time of such a system is

$$T_p=\frac{\pi}{\sqrt{1-\zeta^2}} \tag{2}$$

and the height of the first transient overshoot is

$$m=\exp\left(-\frac{\zeta}{\sqrt{1-\zeta^2}}\pi\right) \tag{3}$$

Consequently, what the timer must do is simply to reduce $r(t)$ in the proportion of $k=1/(1+m)$ for the duration of $T_x=T_p$ after applying the step, and an oscillation-free response of length T_p is obtained. The two responses and inputs without and with the timer are shown in Figs. 2(A) and (B) respectively. Note that the time of reducing the input and the proportion in which it must be reduced are independent of the R_0 height of the step input. Consequently, the timer is simply of the order of a fixed potentiometer switched out

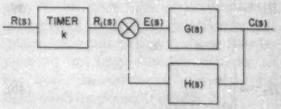


Fig. 1. Sketch of proposed system

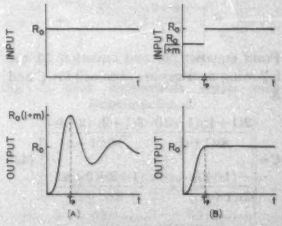


Fig. 2. Qualitative diagrams showing operation of proposed system

A—Without timer
B—With timer

Paper 58-894, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 17, 1958; made available for printing June 3, 1958.

J. ZABORSZKY is with Washington University, and consultant to McDonnell Aircraft Corporation, St. Louis, Mo.

The author wishes to thank Max Wawrziniak for carrying out much of the analog computer documentation for this paper.

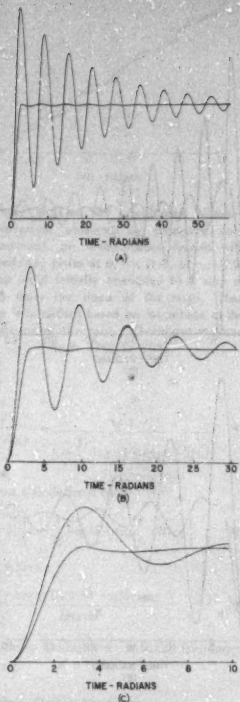


Fig. 3. Unretouched oscillograms showing improvement obtained by proposed nonlinearization on second-order system with closed-loop poles at different values of p_1 and p_2. Time scale is in radians based on magnitude of the complex frequency of the root

A—$p_1 = -0.05 + j0.9982$, $p_2 = -0.05 - j0.9982$. Step input initially reduced in ratio of $k = 0.55$ for duration of $T_s = 3$ radians. Overshoot reduced from 81% to 1%

B—$p_1 = -0.1 + j0.995$, $p_2 = -0.1 - j0.995$. Step input initially reduced in ratio of $k = 0.58$ for duration of $T_s = 3.1$ radians. Overshoot reduced from 70% to 1%

C—$p_1 = -0.3 + j0.9535$, $p_2 = -0.3 - j0.9535$. Step input initially reduced in ratio of $k = 0.740$ for duration of $T_s = 3.67$ radians. Overshoot reduced from 37% to 1%

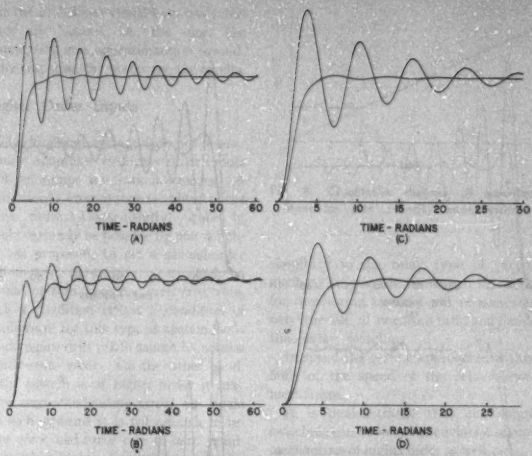


Fig. 4. Unretouched oscillograms showing improvement obtained by proposed nonlinearization on third-order system with closed-loop poles at different values of p_1, p_2, and p_3. Time scale is in radians based on magnitude of the complex frequency of predominant root

A—$p_1 = -0.05 + j0.99875$, $p_2 = -0.05 - j0.99875$, $p_3 = 0.6$. Step input initially reduced in ratio of $k = 0.567$ for duration of $T_s = 3.0$ radians. Overshoot reduced from 35% to 0.5%.

B—$p_1 = -0.05 + j0.9982$, $p_2 = -0.05 - j0.9982$, $p_3 = -0.3$. Step input initially reduced in ratio of $k = 0.633$ for duration of $T_s = 3.0$ radians

C—$p_1 = -0.1 + j0.995$, $p_2 = -0.1 - j0.995$, $p_3 = -0.12$. Step input initially reduced in radio of $k = 0.550$ for duration of $T_s = 3.15$ radians. Overshoot reduced from 50% to 1%

D—$p_1 = -0.1 + j0.995$, $p_2 = -0.1 - j0.995$, $p_3 = -0.606$. Step input initially reduced in radio of $k = 0.534$ for duration of $T_s = 3.0$ radians. Overshoot reduced from 28% to 0.5%

by a clock. No monitoring of either the input, output, or error is necessary.

Analog computer results for second-order systems are shown in Figs. 3(A) through (C) for three different values of damping ζ. These curves show that the results are fully as good as expected. The three curves are on the same real-time scale and they make it very clear that the best way to design the nonlinearized system is to reduce damping ratio ζ as much as practicable. It should be mentioned in this connection that the timer could be used also to raise the damping ratio of the system with the purpose of making any residual oscillations well damped. However, Figs. 3(A) through (C) do not indicate the need for such increased damping in a noise-free environment where these curves were obtained. Theoretically there should be no residual oscillation in a second-order system, but the true order of any physcial system will always be higher than two, so some slight residual oscillation as shown in Figs. 2 and 3, is to be expected.

Comparison With Relay Servomechanism

The proposed system is much superior to the relay-type servos because of the absence of any monitoring circuit and the simplicity of the timer. Of course there is a price to pay for this simplicity; the optimization regarding response time is less complete than the ultimate of optimization represented by the relay type. How much slower is the new system than the relay servomechanism if both are utilizing the same peak torque, in other words, the same servomotor? The relay servo uses peak torque all the time so that without losses its response consists of two parabolic arcs. The ratio of the peak times of the proposed system and the relay-type system is readily calculated as

$$\frac{T_p}{T_{p\ \text{relay}}} = \frac{\pi}{2\sqrt{R_0}} = 1.55\,\frac{1}{\sqrt{R_0}} \qquad (4)$$

where R_0 is the height of the step input.

In other words, the proposed system is slower than the relay-type servo by about 50% for a step input of unit magnitude. In the author's opinion this 50% improvement could seldom justify the additional complications of the monitoring circuit. With viscous damping the ratio is not expressible in closed form, but it can be given as[9]

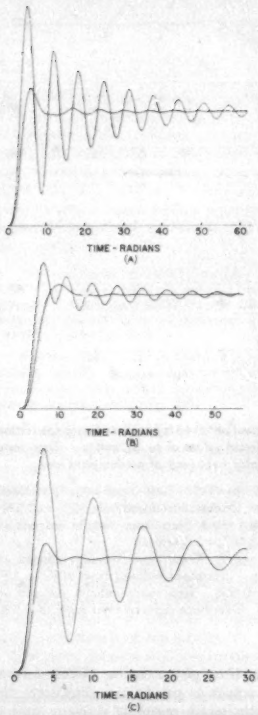
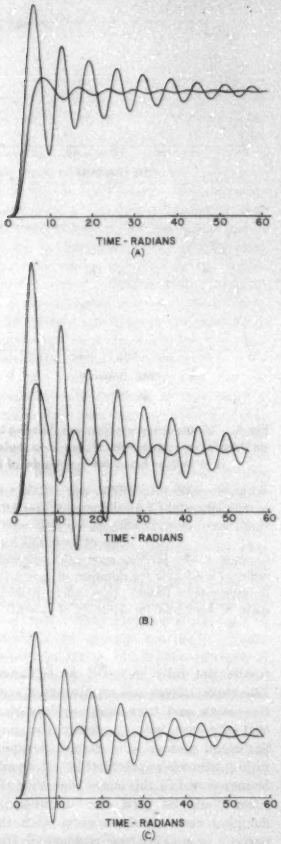


Fig. 5. Unretouched oscillograms showing improvement obtained by proposed nonlinearization on fourth-order system with closed-loop poles at different values of p_1, p_2, p_3, and p_4. Time scale is in radians based on magnitude of the complex frequency of the predominant root

A—$p_1 = -0.05 + j0.9982$, $p_2 = -0.05 - j0.9982$, $p_3 = -0.3 + j0.7145$, $p_4 = -0.3 - j0.7145$. Step input initially reduced in ratio of $k = 0.534$ for duration of $T_s = 3.0$ radians. Overshoot reduced from 94% to 20%

B—$p_1 = -0.05 + j0.9982$, $p_2 = -0.05 - j0.9982$, $p_3 = -0.30 + j0.40$, $p_4 = -0.30 - j0.40$. Step input initially reduced in ratio of $k = 0.534$ for duration of $T_s = 3.0$ radians. Overshoot reduced from 28% to 9%

C—$p_1 = -0.1 + j0.995$, $p_2 = -0.1 - j0.995$, $p_3 = -0.6 + j1.43$, $p_4 = -0.6 - j1.43$. Step input initially reduced in ratio of $k = 0.567$ for duration of $T_s = 3.0$ radians. Overshoot reduced from 94% to 13%



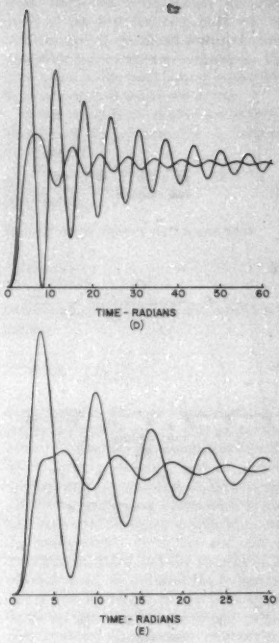


Fig. 6. Unretouched oscillograms showing improvement obtained by proposed nonlinearization on fifth-order system with closed-loop poles at different values of p_1, p_2, p_3, p_4, and p_5, and closed-loop zero at z_1. Time scale is in radians based on magnitude of the complex frequency of the predominant root

A—$p_1 = -0.05 + j0.9982$, $p_2 = -0.05 - j0.9982$, $p_3 = -1.0$, $p_4 = -0.3 + j0.7145$, $p_5 = -0.3 - j0.7145$. Step input initially reduced in ratio of $k = 0.4738$ for duration of $T_s = 3.0$ radians. Overshoot reduced from 71% to 12%

B—$p_1 = -0.05 + j0.9982$, $p_2 = -0.05 - j0.9982$, $p_3 = -1.0$, $p_4 = -0.3 + j0.7145$, $p_5 = -0.3 + j0.7145$, and closed-loop zeros at $z_1 = -0.2$. Step input initially reduced in ratio of $k = 0.472$ for duration of $T_s = 3.10$ radians. Overshoot reduced from 300% to 110%

C—$p_1 = -0.05 + j0.9982$, $p_2 = -0.05 + j0.9982$, $p_3 = -1.0$, $p_4 = -0.3 + j0.7145$, $p_5 = -0.3 - j0.7145$, and closed-loop zeros at $z_1 = -0.4$. Step input initially reduced in ratio of $k = 0.482$ for duration of $T_s = 3.255$ radians. Overshoot reduced from 167% to 42%

D—$p_1 = -0.05 + 0.99875$, $p_2 = -0.05 - 0.99875$, $p_3 = -1.0$, $p_4 = -0.3 + 0.7145$, $p_5 = -0.3 - 0.7145$, and closed-loop zeros at $z_1 = -0.6$. Step input initially reduced in ratio of $k = 0.4707$ for duration of $T_s = 3.285$ radians. Overshoot reduced from 123% to 23%

E—$p_1 = -0.1 + j0.995$, $p_2 = -0.1 - j0.995$, $p_3 = -2.0$, $p_4 = -0.6 + 1.43$, $p_5 = -0.6 - 1.43$, and closed-loop zeros at $z_1 = -0.8$. Step input initially reduced in ratio of $k = 0.492$ for duration of $T_s = 2.9$ radians. Overshoot reduced from 111% to 15%

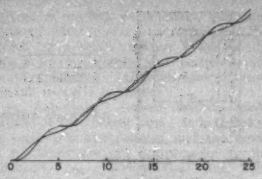


Fig. 7. Unretouched oscillogram showing improvement obtained by proposed nonlinearization on second-order system with closed-loop poles at $p_1 = +j1.0$, $p_2 = -j1.0$. Ramp input initially converted to a step of 0.16 times the slope of the ramp. Time scale is in radians based on magnitude of the frequency of the root. Overshoot reduced from 33% to 20%

$$\frac{T_p}{T_{p\ \text{relay}}} = \frac{\frac{\pi}{\sqrt{1-\zeta^2}}}{t_1 + \frac{1}{2\zeta}\log\left[1+\exp(-2\zeta t_1)\right]} \tag{5}$$

where t_1 is defined by

$$R_0 = 2\zeta\left\{t_1 - \frac{1}{2\zeta}\log\left[1+\exp(-2\zeta t_1)\right]\right\} \tag{6}$$

and where

$$2\zeta = \frac{\text{viscous friction coefficient}}{\text{inertia}} \tag{7}$$

assuming undamped natural frequency of one.

Higher Order Systems

It should be kept in mind that for a third-order system it is not theoretically possible to produce a response of finite duration with only one switching since three quantities i.e., error, error rate, and error acceleration, must be made zero simultaneously for a third-order system to find equilibrium. Nevertheless, Figs. 4 through 6 illustrate that very substantial improvements can be obtained even in fifth-order systems by just one switching. It is true that the transient is not eliminated past the switching time T_s, but the overshoot becomes a small fraction of the unswitched overshoot. It also should be kept in mind that the optimum switching time T_s for higher order systems is not equal any more to the peak time T_p; it tends to be less. Similarly, the best proportion of reduction k, applied during the T_s interval, is not necessarily the value as given previously. The best process of synthesis for higher order systems is probably a cut-and-try method on the analog computer. Usually, the farther from the imaginary axis the additional closed-loop poles and zeros are located, i.e., the more the system response approximates a second-order response, the better are the results.

Higher Order Inputs

The proposed system does not appear readily adaptable to higher order inputs such as ramps for various reasons. A zero velocity error system is at least of the third order. Consequently, while it would certainly be possible by one switching, as proposed, to get a second-order system with a ramp input to reach an instant of zero error and zero error rate, such a condition is not a condition of equilibrium for this type of system, so it would again drift off to asume its normal steady-state error. On the other hand, if the system is of higher order it may have zero steady-state error, but again for such systems it is not enough to reduce error and error rate to zero simultaneously; higher derivatives must also be zero and this is unobtainable by just one switching. Nevertheless some improvement like the one demonstrated in Figs. 4 through 6 could be expected and the best parameters should be established by analog computer work. A single example is shown for this in Fig. 7 for the simplest possible zero velocity error system. The switching circuit used here changed the initial part of the ramp input to a step, as sketched in Fig. 8. T_s here would be a constant so the height of the step would be proportional to the slope of the ramp.

Noise

As described here the system is proposed for a noise-free environment. Noise, however, should not eliminate the usefulness of the system since the difference with noise is mostly that the value reached at peak time is off by some random deviation depending on the noise. Thus, instead of remaining constant after the switching the output will now undergo some residual oscillations which, are however, still much smaller than the oscillations without the timer. In the presence of noise it will be almost imperative to change the system damping also at the switching. This raises the question of optimum damping before and after the switching, a rather involved problem which is now under study.

Conclusions

A nonlinear or sectionally linear control system is discussed which is superior in simplicity to the relay type of servomechanism, since it eliminates the need for monitoring circuits and necessitates only the use of one fixed ratio and fixed-time-delay switching.

In speed the system approaches within 50% of the speed of the relay servomechanism.

It is demonstrated that one single switching can be very beneficial for servomechanisms of higher order as well.

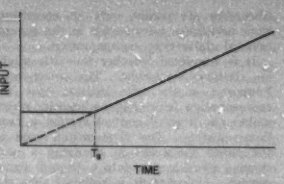


Fig. 8. Qualitative diagram of modified input for zero velocity error system

References

1. Poincast Control of Damped Oscillatory Systems, O. J. M. Smith. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, Sept. 1957, pp. 1249–55.

2. Analog Study of Dead Beat Poincast Control, G. H. Tallman, O. J. M. Smith. *Transactions*, Professional Group on Automatic Controls, Institute of Radio Engineers, vol. PGAC-4, Mar. 1958, pp. 14–21.

3. Feedback Control Systems (book), O. J. M. Smith. McGraw-Hill Book Company, Inc., New York, N. Y., 1958, pp. 331–45.

4. Signal Component Control, D. J. Gimpel, J. F. Calvert. *AIEE Transactions*, vol. 71, pt. II, Nov. 1952, pp. 339–43.

5. The Design of Transversal Filters Using Time Domain Methods, J. A. M. Lyon, T. B. Thompson. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, p. 532.

6. The Application of Short-Time Memory Devices to Compensator Design, D. J. Ford, J. F. Calvert. *AIEE Transactions*, vol. 78, pt. II, May 1954, pp. 88–93.

7. Extended Switching Criterion for Second-Order Saturated Servomechanisms, J. W. Diesel. *Ibid.*, vol. 76, pt. II, 1957 (Jan. 1958 section), pp. 388–93.

8. A Phase-Plane Approach to the Design of Saturating Servomechanisms, A. M. Hopkin. *Ibid.*, vol. 70, pt. I, 1951, pp. 631–39.

9. Effects of Friction in an Optimum Relay Servomechanism, T. M. Stout. *Ibid.*, vol. 72, pt. II, Nov. 1953, pp. 329–36.

10. Short-Time Memory Devices in Closed Loop System—Steady-State Response, T. W. Sze, J. F. Calvert. *Ibid.*, vol. 74, pt. II, Nov. 1955, pp. 340–44.

Discussion

D. J. Ford and T. W. Sze (University of Pittsburgh, Pittsburgh, Pa.): The author is to be commended for presenting a rather interesting way of instrumenting a delay-type compensation for a linear second-order system. The comments prepared by the

discussers are meant only to contribute some additional information in the general area of delay-type compensation.

It is wondered whether the title might be clarified somewhat. The word nonlinearized in the title implied that the compensatiag device has nonlinearity which makes the over-all control system a nonlinear system. The time switch used in the paper succeeded in changing a continuous signal system into a discrete signal system. However, it is believed that the introduction of the timer switch did not bring in any nonlinearity into the system component, and hence the system remained linear. This is verified by the fact that the principle of superposition of linear analysis is still implied in the paper, as was indicated by the author's statement in the last paragraph under the section entitled "How Nonlinearities Can Optimize the Response." Furthermore, a fixed-time-delay operator can be considered, by Taylor's series expansion, as an infinite series of linear operators of increasing orders. Consequently, the effect of a timer switch upon the system transfer function is that it adds an infinite number of zeros to the system, and yet the poles are not disturbed. The system remains a linear one.[1]

Dr. Zaborszky makes mention of the improved response attainable with the on-off relay-controlled saturating servo studied by Diesel[2] in 1957 and earlier by Hopkins[3] in 1951. It might be pointed out that the controlled systems considered by Dr. Zaborszky are not the nonlinear saturating types treated by Hopkins and Diesel. The systems considered in the paper are rather the strictly linear types of second and higher order systems. The author's work and Diesel's and Hopkins' works should probably, then, be compared with considerable caution since they are designed for use with two different types of controlled systems. It is well recognized that the performance of nonlinear controlled systems frequently differs greatly from that of strictly linear controlled systems.

In the Gimpel-Calvert paper,[4] the design of delay-type compensator was carried out on a completely rigorous analytical basis. The solution for the various compensator parameter values appear in closed form so that it is not necessary to utilize any analog computer trial-and-error procedure for their determination. Further, it is believed that the Gimpel-Calvert development gave the design equations for the complete compensation of an nth order linear system subjected to an mth degree polynomial time function input, where n and m can be any real positive integers including zero. These higher order and higher degree cases do require timers with additional stops, or delay lines with additional taps. It would appear, then, that Dr. Zaborszky's system represents a special case of the Gimpel-Calvert theory. The timer-potentiometer mechanism suggested in this paper does provide a very novel and very simple way of instrumenting this type of compensator.

To illustrate the relationship between the two papers, the following general development of the Gimpel-Calvert paper is presented for the second-order case considered by Dr. Zaborszky.

Fig. 9 shows Zaborszky's proposed system redrawn in terms of a memory-device compensator. Referring to the Gimpel-Calvert paper, we find that this system is a special case of design equations 8(A), 8(B), and 8(C), since B_3, B_4, etc., are zeros.

The design equations for the compensator are then:

$$\left.\begin{aligned} &B_0+B_1=1\\ &B_0+B_1e^{-\mu_1 T_x}=0\\ &B_0+B_1e^{-\mu_2 T_x}=0 \end{aligned}\right\} \tag{8}$$

where μ_1 and μ_2 are the poles of the controlled second-order system, and

$$\mu_1=-\zeta-j\sqrt{1-\zeta^2}$$

$$\mu_2=-\zeta+j\sqrt{1-\zeta^2}$$

If these equations are solved simultaneously, it can be shown that:

$$\left.\begin{aligned} T_x&=\frac{n\pi}{\sqrt{1-\zeta^2}}\\ B_0&=\frac{(-1)^{n+1}e^{\zeta T_x}}{1+(-1)^{n+1}e^{2\zeta T_x}}\\ B_1&=\frac{1}{1+(-1)^{n+1}e^{\zeta T_x}} \end{aligned}\right\} \tag{9}$$

where n is any positive integer. Thus, k (Zaborszky paper) = ratio of magnitudes of the step function before and after switching

$$k=\frac{B_0}{B_0+B_1}=B_0 \tag{10}$$

Equation 8 gives the general design conditions of a memory-device compensator to reduce the error and the error rate to zero after and including the nth overshoot (or undershoot) of the transient response of the uncompensated system. This is because the first condition of equation 10 is recognized as the time instant when the nth overshoot or undershoot peak occurs. Thus, if we wish to eliminate the error and the error rate at and after the first transient overshoot, we would select $n=1$. Equation 8 thus yields for a special case of $n=1$, the following set of design conditions for the compensator:

$$\left.\begin{aligned} T_x&=\frac{\pi}{\sqrt{1-\zeta^2}}=T_p\\ B_0&=\frac{e^{\zeta T_p}}{1+e^{\zeta T_p}}=\frac{m^{-1}}{1+m^{-1}}=\frac{1}{1+m}\\ B_1&=\frac{1}{1+e^{\zeta T_p}}=\frac{1}{1+m^{-1}}=\frac{m}{1+m}\\ k&=B_0=\frac{1}{1+m} \end{aligned}\right\} \tag{11}$$

where m is defined in Dr. Zaborszky's paper. Thus are the design conditions used in Dr. Zaborszky's paper. These procedures illustrated here are not intended to show the superiority of the Gimpel-Calvert method of designing a compensator for a simple second-order system. As a matter of fact, the author used a much simpler reasoning process to reach the same answer. However, these procedures demonstrated here may be applied to the design of compensators for a high-order system and high-degree input time function. In such a case, as mentioned before, a multiswitching timer, or a delay line with many taps, or its equivalent, should be used. The algebraic work will be increased, but there is no increase in conceptional difficulty.

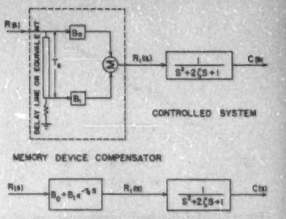


Fig. 9. Zaborszky's proposed system redrawn as memory-device compensated system

In the section entitled "The Proposed System," there seems to be the implication that this type of time-delay compensator works best for the low values of damping. We are in complete agreement with this from the point of view of hardware or instrumentation. As is seen by the first section of equation 8 or equation 11, the low values of damping do produce the smaller required time delays, T_x or T_p. In terms of instrumentation, these smaller values are desirable. However, with the design theory presented by Gimpel and Calvert, complete compensation is possible for any value of the damping factor less than unity. If the damping factor is unity or greater, the system response may overshoot, but it will contain no oscillatory terms. In that case, neither the Zaborszky theory nor the Gimpel-Calvert theory is applicable.

References

1. Transients in Linear Systems (book), M. F. Gardner, J. L. Barnes. John Wiley & Sons, Inc., New York, N. Y., 1942, pp. 236–41.
2. See reference 7 of the paper.
3. See reference 8 of the paper.
4. See reference 4 of the paper.

Otto J. M. Smith (University of California, Berkeley, Calif.): Dr. Zaborszky has done an excellent laboratory study of a half-cycle Posicast deadbeat control system. He is to be complimented on the extensive tests which he has performed in verifying the fundamental theory of this system. Reference 3 of the paper describes in detail how the residual oscillations in Fig. 6(B) can be reduced through the use of additional delayed actions. In every case, the use of two timers and two auxiliary relays can significantly increase the speed of action without increasing the power required. In the half-cycle or half-period control of Dr. Zaborszky, only half power is applied for one half of the cycle and then full power following that time. In one-third period control, full power can be applied immediately, removed after one-sixth period and reapplied after one-third period to achieve a more rapid rate of rise of the output and no oscillations. For only a slight increase in power, one-quarter period control can be achieved. In this case a slight excessive

voltage is applied immediately, and then removed after one-eighth period and replaced by a small negative value for the time from one eighth of a period to one fourth of a period. After one-fourth period, the full input is applied.

Dr. Zaborszky's system uses the timer outside of the control loop. It is therefore an open-loop compensator and the errors in the timer are not compensated for by feedback. More serious than this, however, is the fact that load disturbances also do not go through the timer. Changes in required torques and other types of load disturbances which the system would have in the absence of the timer should be compensated for by the timer. In a practical system, therefore, it is not feasible to leave the timer outside of the loop.

Techniques for incorporating the timer into the feedback loop, so that all load disturbances will produce deadbeat transients and will not excite oscillations, are shown in Fig. 6 of reference 1 and Fig. 9 of reference 2 of this discussion, and also in Fig. 13 of reference 1, Figs. 7 through 9 of reference 2, and Figs. 10 through 29 of reference 3 of the paper. I shall look forward with great interest to additional publications of Dr. Zaborszky's on this subject. He is to be commended highly for his work.

References

1. Direct Synthesis Through Block Diagram Substitutions, Otto J. M. Smith. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1957, pp. 8–16.

2. Mixed Distributed and Lumped Parameter Systems, Otto J. M. Smith. *IRE WESCON Convention Record, Ibid.*, pt. 2, 1957, pp. 122–32.

J. Zaborszky: The author appreciates the contribution of the discussers since these not only add further pertinent information, but also provide the author with an opportunity to elaborate on some aspects of the proposed system, notably its relative position with respect to other similar systems. Generally speaking the Posicast control system of O. J. M. Smith,[1] the short memory compensator system of J. F. Calvert,[2] and the system discussed in this paper are mutually special cases of each other, or nearly so, for the simplest case of a second-order system and one switching. However, in contrast to the other two, the system discussed in this paper utilizes just one switching regardless of the order, and the paper demonstrates how much improvement can be obtained by the simple means applied. In fact, the entire emphasis of the paper is on a striving for simplicity and the accompanying small weight and low price. While it is possible (and the higher refinement in performance may make it very worth while) to instrument a co-ordinated multiple-switching circuit or delay line, the amount of gadgetry, circuitry, weight, and price can become quite prohibitive at least in some time ranges and at certain locations.

The discussion of D. J. Ford and T. W. Sze is first treated in a point-by-point fashion. The first point is in their first sentence which expresses an impression that the paper is being devoted to second-order systems. Actually, most of the work reported in the paper concerns higher order systems. Eleven out of a total of 14 oscillograms refer to higher order systems.

The second point of the discussers is that the system proposed in the paper is linear. The author is forced to disagree strongly. Consider the superposition of two step inputs with the second one arriving, for instance, while the switch is still open for the first. Obviously the superposition principle does not apply, the joint response will not be the sum of the responses for the separately applied inputs, because either the switch will stay open too long for the first input or it will close too soon for the second, depending on the specific instrumentation. (With certain types of instrumentation the closing time may even be correct for both inputs in which special case the system reduces to a linear one.) Consequently time-delay operators do not apply in general to the system proposed in the paper since Laplace transformation does not apply to nonlinear differential equations. It appears to be implied in the discussion that to keep a system linear one could add an infinite number of zeros but not poles, actually one loses the entire s-plane representation, poles, zeros, and all else for a nonlinear system such as the one proposed in the paper. The reason the system proposed in the paper looks linear to the reviewers seems to be that they limit their thinking to delay-line type of instrumentations. If so, however, they are only discussing one possible special case of the paper, which special case is far from the most desirable in many applications.

The third point of the discussers, apparently through an oversight, voices caution about comparing the author's work with that of A. W. Hopkins[3] and J. W. Diesel.[4] No such comparison is implied in the paper. The paper is merely quoting the proofs given by these authors, especially Diesel, to the effect that a relay-type servomechanism represents the maximum optimization of the response time which is attainable with given physical equipment. The paper does contain a comparison of the proposed system with the relay-type servomechanism which of course has been studied by a great many authors beside the afore-mentioned, in fact the work of one, T. M. Stout,[5] is actually quoted in this comparison. The author could not agree more that this comparison, and any comparison of different systems for that matter, should be made with utmost care and realism. The author is firmly convinced however that the comparison given in the paper is just such a careful and realistic comparison. In the paper the relay-type servomechanism and the second-order case of the proposed system are compared on the basis of equal inertia, equal damping, and equal maximum servomotor torque, or in other words equal servomotor size and equal load; this means the same plant or physical equipment excluding compensating, monitoring, and switching circuitry. This appears to be the most logical and realistic basis for comparison since it should yield the answer to the problem of the designer: i.e., how fast can a given load be positioned by a given servomotor depending on the selection of a relay-type circuitry or the circuitry proposed in the paper. This is a clear match of efficiency versus simplicity.

The fourth point of the discussers is a demonstration of computing the switching parameters for a second-order system by the equations of Gimpel and Calvert.[6] The discussers themselves sight the fact that this process is much more cumbersome than the trivially simple derivation in the paper, but correctly claim that this complexity is necessary for designing the short memory compensator of Calvert for higher order systems and inputs. The important point here is, that, as the discussers themselves state, this design would lead to multiple switching for higher orders and consequently it would design for higher orders a system totally different from that discussed in the paper.

The fifth and last point of the discussers sights an implication in the paper to the effect that the proposed system works best for low values of damping. There is no such implication in the paper; of course the system works equally well for any damping $\zeta<1$. However, naturally, more is to be gained by introducing switching if the damping is small since for such systems overshoot is high and peak time tends to be short. Consequently if the designer has freedom of choice he may wish to replace a well-damped system without switching by a slightly damped system with switching. This would permit him to compound his gains derived from introducing switching.

In connection with Professor Smith's discussion, the first point again is that for the second-order system, with one switching, the Posicast control as described in reference 1 of the paper and the system discussed in this paper are mutually special cases of each other. The author too wishes to call attention at this time to the theoretical work in reference 3 of the paper which along with reference 2 of the paper was published about the time when this paper was submitted to AIEE so that references 2 and 3 could only be added later in the final printed form of the paper. In fact the first report on the material of this paper was made in August 1957, before the publication of reference 1.

The author feels compelled strongly to disagree with the contention of the second point. The realistic way to compare two systems of switching compensators, is to assume the same plant or physical equipment, i.e., the same inertia load and the same servomotor with a given peak torque, and then see how fast this load can be positioned depending on the selection of the control or switching circuits. In the system discussed in this paper the full forward torque of the servomotor is used initially with the nominal step imput, and no higher torque, forward or reverse, is applied thereafter. This means that, unless full saturated torque is always maintained by the switching circuit as in the relay-type servomechanism, no faster response is possible with the given servomotor than that of the system discussed in the paper. In other words it would actually necessitate replacing the servomotor with one of higher torque capabilities to produce 1/3- or 1/4-cycle response; however, as is explained in the paper, the application of a relay-type switching circuit will permit about 30% reduction of response time with the same servomotor at the expense of much more complex monitoring and switching circuitry.

Prof. Smith's remarks regarding closed-loop control are very pertinent for cases where delay-line type of instrumentation

is used, since such devices do not have a transfer function of one or a gain of one. Such instrumentations are not however primarily visualized for the system of this paper. What is intended here is a relay which reduces the input by a certain percentage for a given length of time and produces a direct connection thereafter. Direct connections have inherently transfer function one, thus no closed loop is needed here. Any inaccuracies in either the timing or the initial percentage reduction will only cause small residual oscillations. In instrument-type servomechanisms, which are the subject of this paper, there are no torque loads, and no load changes. Even in most other systems load changes produce residual oscillations, i.e., oscillations of much smaller amplitude than those caused by the displacement inputs. Consequently it will probably seldom be justified to go to the more elaborate closed-loop devices mentioned by Prof. Smith.

Again it should be emphasized that theoretically more perfect compensation can be obtained by more complex compensating circuits. The aim of this paper is however not the ultimate refinement in compensation, but a demonstration of how much improvement can be obtained by very simple tools which will repay in small weight, size, price, and increased reliability for some reduction in the quality of performance.

REFERENCES

1. See reference 1 of the paper.
2. See reference 10 of the paper.
3. See reference 8 of the paper.
4. See reference 7 of the paper.
5. See reference 9 of the paper.
6. See reference 4 of the paper.

Frequency Response of Nonlinear Closed-Loop Feedback Control Systems

S. L. MIKHAIL
MEMBER AIEE

G. H. FETT
MEMBER AIEE

THE ANALYSIS of nonlinear control systems is not a new problem. In 1918, Barbillion[1,2] investigated the use of a hydraulic servomotor in a nonlinear regulating control system. The work of Babillion and Cayere is a modification of the principle given by Leaute' in 1885 in his treatise on the prevention of long period oscillations in machines driven by water turbines. In considering the operation of the servomotor, Leaute' suggested that the servomotor should be advanced to damp out the oscillations in question. In all cases there is more or less considerable oscillation of speed and the regulation is made by successive approximations.

In recent literature, considerable attention has been given to sinusoidal analysis of nonlinear systems. Goldfarb,[3] Tustin,[4] Kochenburger,[5] and Johnson[6] adopted the method of representing the nonlinear system by a describing function. The latter is defined as the ratio of the fundamental sinusoidal component of the output to the input sine wave. Most of the use of the sinusoidal analysis of nonlinear systems has been in the direction of using a describing function that is dependent only on the amplitude of the input.[7,8] Very little attention has been given to the analysis of closed-loop servo systems using describing functions that are both amplitude- and frequency-dependent. El-Sabbagh[9] and Stein and Thaler[10] have started some work in this direction. El-Sabbagh found the frequency response of a closed-loop system having a nonlinear and a linear element for a constant output of the nonlinear element. Stein and Thaler discussed a closed-loop nonlinear system in the gain-phase plane; they used a trial-and-error method and employed the Nichols' chart. A discussion[11] by one of the authors correlates these two references. Stein and Thaler used a 2-phase servomotor[10,12,13] as their nonlinear element that is both amplitude- and frequency-dependent.

It is the purpose of this paper to present a graphical method for the determination of the frequency response of a closed-loop servomechanism with two nonlinearities in series. The describing functions are dependent on the frequency and amplitude of the output of one or the input to the other. One of the nonlinearities may be in the feedback path. Jump phenomena can be observed and system stability can be determined from the construction. In a subsequent paper it is planned to discuss some applications of the proposed method. Simple as it is, the method of dual loci presented here is believed to be novel.

Paper 58-895, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 17, 1958; made available for printing April 17, 1958.

S. L. MIKHAIL is with the University of California, Berkeley, Calif., and G. H. FETT is with the University of Illinois, Urbana, Ill.

The authors wish to acknowledge the invaluable discussions and comments of Professors O. J. M. Smith, A. M. Hopkin, and E. I. Jury of the University of California, Berkeley.

Transfer Functions

With reference to Fig. 1, let G_L be a linear function, i.e., its input and output are related by a linear differential equation with constant coefficients. If the input is a sine wave $A_i \sin 2\pi ft$, then the output will also be a sine wave of the same frequency that will be designated as $A_o \sin (2\pi ft+\phi)$. The transfer function G_L is thus a phasor operator of magnitude A_o/A_i and has an angle ϕ. It thus can be represented by a phasor operator in the complex plane as shown in Fig. 2. Since the function is linear, if the frequency is kept constant G_L will not vary with the variation of the amplitude of the input. If, however, the frequency is varied, a different value is in general obtained for G_L for every frequency. If the ends of all the phasor operators representing G_L at various frequencies are joined, a transfer function locus is obtained as in Fig. 2.

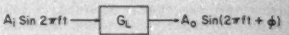


Fig. 1. Transfer function of a linear relation

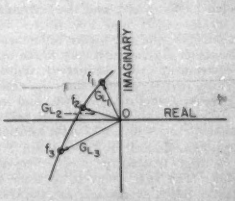


Fig. 2. Representation of a transfer function G_L in the complex plane

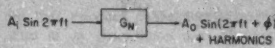


Fig. 3. Describing function of a nonlinear relation

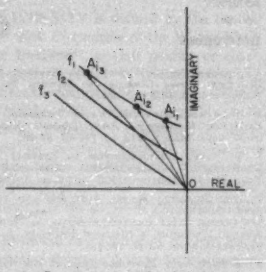


Fig. 4. Representation of a describing function G_N in the complex plane

Describing Functions

Now let the input and the output be related by a nonlinear differential equation. Thus if a sinusoidal input $A_i \sin 2\pi ft$ is assumed, the output will in general contain a fundamental $A_o \sin(2\pi ft+\phi)$ plus harmonics. The describing function G_N is defined as a phasor operator that, if multiplied by the phasor for the sinusoidal input phasor, gives the phasor for the fundamental component of the output. Thus in this case $G_N = A_o/A_i\underline{/\phi}$ and can thus be represented by a phasor operator in the complex plane. Keeping the frequency constant and varying the amplitude A_i of the input would in general vary G_N because of the nonlinearity. By joining the ends of the phasor operators obtained for a constant frequency and variable input amplitude, a curve is obtained. Repeating the same process for different constant frequencies f_2, f_3 . . . and a variable input amplitude, G_N can be represented by a family of curves of constant frequencies each scaled in terms of the amplitude of the input A_i. It is worth mentioning that with certain types of nonlinearities, for a constant frequency, G_N may be a multivalued function of the amplitude and thus more than one point on a constant frequency curve may carry the same index for amplitude. It is also obvious that, since the input and G_N specify the fundamental of the output, the curves of constant frequencies of G_N can very well be scaled in terms of the amplitude of the fundamental of the output.

In general, a nonlinear function may include both linear and nonlinear elements. See Figs. 3 and 4.

Closed-Loop Servomechanisms with Two Nonlinearities

In the system of Fig 5 let R be the sinusoidal input and E and C the fundamental components of the error and controlled quantities, respectively. Now G_A and G_B are two nonlinearities that are given by their describing functions and are dependent upon the amplitude and frequency of the fundamental of N, which is the output of G_A and the input to G_B. From Fig. 5,

$$R = E + C \tag{1}$$

Since R, C, and E represent sine waves, equation 1 is a phasor equation and the addition used in it is phasor addition. Division of equation 1 by the phasor N gives

$$R/N = E/N + C/N \tag{2}$$

$$= 1/G_A + G_B \tag{3}$$

Equations 2 and 3 are phasor operator equations and can be represented graphically in the complex plane as in Fig. 6.

Draw the curves representing $-1/G_A$ and G_B as in Fig. 6. These curves will be constant frequency curves scaled by the amplitude of N. Any $-1/G_A$ curve carrying the same frequency index as a G_B curve corresponds to this curve, e.g., the two curves having frequencies f_1 in Fig. 1 correspond to each

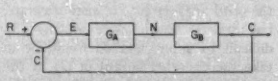


Fig. 5. Closed-loop servomechanism with two nonlinearities

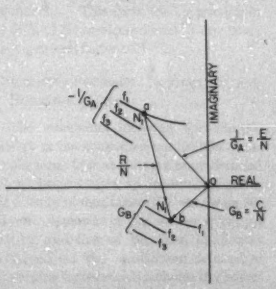


Fig. 6. Phasor operator representation of E/N, C/N, R/N

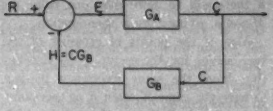


Fig. 7. Closed-loop servomechanism with nonlinearity in the feedback path

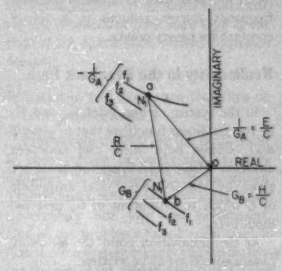


Fig. 8. Phasor operator representation of $1/G_A$, G_B, and R/C

other. On these two curves, points having the same index for N correspond to each other. In Fig. 6, points a and b thus correspond to each other. From Fig. 6 it is seen that

$$\overline{ao} = 1/G_A = E/N \tag{4}$$

and

$$\overline{ob} = G_B = C/N \tag{5}$$

Thus

$$\overline{ab} = R/N \tag{6}$$

giving

$$M = |C/R| = \overline{ob}/\overline{ba} \tag{7}$$

and the angle ϕ between C and R is given by

$$\phi = \underline{/abo} \tag{8}$$

Thus if M and ϕ are desired for any frequency f_1 and any value of the input R_1 the following steps are to be followed:

The family of curves for $-1/G_A$, G_B is drawn. The two curves for constant f_1 on $-1/G_A$, G_B are chosen. Since any two points having the same index for N on these two curves have a distance between them equal to R/N these curves can be scaled in terms of R. Thus on these two curves the points a and b are located that have the same R_1 and equations 7 and 8 give M, ϕ at f_1 and R_1.

Obviously, if either $-1/G_A$ or G_B or both are multivalued in terms of the amplitude of N they will also be multivalued in R and thus for the same f and

R there will be more than one value for M and ϕ, giving a jump phenomenon.

Also, if the points of crossing of the two curves for the $-1/G_A$, G_B families that have the same frequency happen to have the same N, then R at this point is zero and C/R is unbounded which indicates a point of instability.

Obviously, if either G_A or G_B is a linear function, then its family of constant frequency curves reduces to a set of constant frequency points.

Nonlinearity in the Feedback Path

In the system of Fig. 7 let G_A and G_B be two nonlinear functions that in general may contain both linear and nonlinear elements whose describing functions are given by the frequency and amplitude of C which is the output of G_A and the input to G_B.

At the summation point the following phasor equation can be written:

$$R = E + H$$

$$= E + CG_B \tag{9}$$

Division of equation 9 by C gives the following phasor operator equation:

$$R/C = E/C + G_B \tag{10}$$

or

$$R/C = 1/G_A + G_B \tag{11}$$

In Fig. 8 the families of curves of $-1/G_A$ and G_B are represented as constant frequency curves scaled for various magnitudes of C. For any curve of the $-1/G_A$ family drawn for a constant frequency f_1 there corresponds another curve in the family G_B also drawn for a frequency f_1. Any two points a,b on these two curves having the same value for C correspond to each other. Thus

$$\overline{ao} = 1/G_A; \quad \overline{ob} = G_B$$

and according to equation 11

$$\overline{ab} = R/C \tag{12}$$

Thus

$$|R/C| = |1/M| = \overline{ab} \tag{13}$$

and the angle that b a makes with the real axis is ϕ.

Thus

$$M = |C/R| = 1/\overline{ab} \tag{14}$$

and ϕ = the angle that b a makes with the positive real axis.

Since a and b carry the same index for C and $\overline{ab} = R/C$ the value for R can be calculated at a and b and thus the curves for $-1/G_A$ and G_B can be scaled in both C and R.

The statements previously made regarding jump phenomena and system stability apply here also. Also, if either G_A or G_B is a linear function its family of constant frequency curves will be replaced by a family of constant frequency points.

The describing function method assumes that the input to the nonlinear element is a sinusoidal function of time, even though the signal actually present may contain harmonics fed back through the feedback element. Harmonics are assumed to have been attenuated by the linear part of the system so as to be negligible.

The method described here for two nonlinearities has the same limitation. For example, if the first nonlinearity G_A is characterized by saturation and the second, G_B by backlash, the response may well be different from that when the order is reversed. A more accurate, but ever so much more tedious and difficult to apply, method would be to take the distorted output of G_A as the input to G_B and determine the transfer function. Hence the answer obtained by the method of this paper is only an approximation which supposes that the input to each nonlinearity is, in turn, a sine wave. When the degree of nonlinearity is small the answer is quite accurate. When the nonlinearity is great the harmonics strongly affect the accuracy.

References

1. Regulation of Generating Systems, L. Barbillion. *Revue Générale des Sciences*, Paris, France, vol. 29, 1918, pp. 469–74.

2. New Regulator for Water Turbines, L. Barbillion, P. Cayere. *Revue Générale d'Electricité*, Paris, France, 1918, pp. 747–64, 87–96.

3. On Some Nonlinearities in Systems of Regulations, L. C. Goldfarb. *Avtomatika i Telemekhanika*, Moscow, USSR, vol. 8, no. 5, 1947.

4. The Effects of the Backlash and the Speed-Dependent Friction on the Stability of the Closed Cycle Control Systems, A. Tustin. *Journal*, Institution of Electrical Engineers, London, England, vol. 92, pt. 2A, 1947.

5. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, R. J. Kochenburger. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 270–84.

6. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements, E. C. Johnson. *Ibid.*, vol. 71, pt. II, July 1952, pp. 169–81.

7. Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities, H. D. Greif. *Ibid.*, vol. 72, pt. II, Sept. 1953, pp. 243–48.

8. A Generalized Method for Determining the Closed-Loop Response of Nonlinear Systems, L. T. Prince. *Ibid.*, vol. 73, pt. II, Sept. 1954, pp. 217–24.

9. A General Method for Analyzing and Synthesizing the Closed Loop Response of a Linear and a Nonlinear Servomechanism, Hassan H. El-Sabbagh. *IRE WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1957, pp. 58–70.

10. Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-Sensitive Describing Function, W. A. Stein, G. J. Thaler. *AIEE Transactions*, vol. 77, pt. II, May 1958, pp. 91–96.

11. *Discussion* by S. L. Mikhail of reference 10, p. 96.

12. *Discussion* by G. H. Fett of reference 10, p. 96.

13. Transfer Function of 2-Phase Servomotors, S. L. Mikhail, G. H. Fett. *Ibid.*, pp. 97–98.

14. Feedback Control Systems (book), G. H. Fett. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1954, p. 333.

Diversity of Farm Loads

LANDY B. ALTMAN, JR.
NONMEMBER AIEE

LEON F. CHARITY
NONMEMBER AIEE

DIVERSITY is defined by the Statistical Committee of the Edison Electric Institute[1] "as that quality or characteristic by which individual maximum demands occur at different times. There is hourly, daily, weekly, monthly, and annual diversity." Obviously, if there were no diversity between appliances, homes, farms, machines, and manufacturing plants the electrical industry would be required to have considerably more capacity than at present.

Electrical characteristics of farm loads including diversity deserve more attention and study. More information would be useful to leaders of the electrical industry who now recognize that the load growth potential on farms is high. Often energy consumption is the only known characteristic upon which to base estimations of needed generation, transmission, and distribution facilities. Also of importance is the lack of information on which to base programs designed to promote desirable types of loads.

To provide some of this needed information, the Farm Electrification Section, Agricultural Research Service, U. S. Department of Agriculture, in co-operation with the Agricultural Engineering Department and the Statistical Laboratory of the Iowa Agricultural and Home Economics Experiment Station and many of the power suppliers in Iowa, has made a number of studies[2,3] of some of the electrical characteristics of Iowa farm loads. These studies were based on data obtained through metering of farms and appliances.

The service entrances to 16 farms were metered one week out of each month for a year with recording ammeters and voltmeters. Forty-two other farms, each having a range, were metered for similar periods with recording demand meters. In another study, 36 highly electrified farms and 9 of the homes were metered continuously for one year with recording demand meters. Recording meters were also used to record when certain appliances were operated and to measure their demands in watts or amperes. Included among the appliances were bulk milk coolers, motors on water pumps and elevators, poultry brooders, heat lamps for pig brooding, automatic stock waterers, and electrically heated homes.

Terms appearing in this paper that may require definitions for readers not closely associated with this subject are as follows:

Demand factor is the ratio of the maximum demand of a system, or part of a system, to the total connected load of the system under consideration.[4]

Maximum demand is the greatest demand of a particular type of load occurring within a specified period.[4]

Load factor is the ratio of the average load over a designated period to the peak load occurring in that period.[4]

Diversity factor is the ratio of the sum of noncoincident maximum demands of two or more loads to their coincident maximum demands for the same period.[1]

Coincident demand is any demand that occurs simultaneously with any other demand; also the sum of any set of coincident demands.[4]

Diversity on Individual Farms

Only part of the connected electrical load on a farm is in use at one time. In the previously mentioned studies of 36 and 42 farms, the 30-minute maximum demand of the average farm was 22% of the connected load, in this case equivalent to a diversity factor of 4.5. Values ranged from 13 to 34% except for one dairy farm with a minimum of household electric equipment. This farm used a maximum of 54% of its connected load at one time, i.e. a demand factor of 54%.

Diversity Between Production and Household Uses

The maximum demand for electric energy in the home does not usually occur at the same time as the maximum demand of appliances used in farm production. In a study of nine farms the maximum 30-minute demands of the homes averaged 6.0 kw and that of the farm appliances averaged 4.1 kw. Maximum demand of these nine farms which includes the homes averaged only 7.1 kw, giving a diversity factor of 1.42. The load factors for farm appliances average 14% while that of the house averaged 15%. Over-all farm load factors averaged 21%. There is also diversity in time of electric use among the farm buildings—the shop, farrowing house, milk house, barn, poultry house, and granary. Diversity should be considered in the choice of transformer size and service and branch feeder conductor sizes.

Even though diversity existed between the house and farm production appliances, electric demands of both were high at the time of the distribution system maximum demand.

Estimating the Maximum Demands of Individual Farms

Transformers which are too small result in poor voltage to farms and excessive numbers of burned-out transformers. Transformers which are too large cause unnecessary line loss and investment. One of the problems facing rural electric distribution system management is matching the size of the distribution transformer to the maximum demand of the load served. The solution to this problem is to develop a method whereby the demand of an individual consumer can be accurately estimated.

Attempts are now being made at Iowa State College to derive equations which will associate factors contributing to the maximum demands of farms. It is hoped that the procedures developed may be used to derive other expressions for estimating the maximum demand of any farm and therefore the size of the transformer. One such equation was obtained from data on 36 highly electrified farms. These farms were among the upper 10% in energy consumption by Iowa farms. Average energy used was 13,854 kwhr (kilowatt-hours) per year. The low farm used 7,112 and the high farm 33,093 kwhr per year. The equation obtained is as follows:

$$Y = 3.398 + 1.233X_1 - 0.282X_3 + 2.096X_4 - 0.845X_4 + 0.857X_5 + 0.529X_6 + 0.00213X_7 + 5.714X_8 - 0.0321X_9 + 0.001799X_{10}$$

where

Y = the predicted annual 30-minute maximum demand in kilowatts of an individual farm
X_1 = number of ranges
X_3 = number of water heaters
X_3 = number of clothes dryers
X_4 = number of freezers
X_5 = number of dairy water heaters
X_6 = number of crop dryers
X_7 = number of stock waterers
X_8 = number of feed grinders
X_9 = number of heat lamps
X_{10} = number of kilowatt-hours used in the month with highest energy consumption

Paper 57-1159, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Farm Electrification Conference, Minneapolis, Minn., October 29–31, 1957. Manuscript submitted September 6, 1957; made available for printing March 7, 1958.

LANDY B. ALTMAN, JR., is with the U. S. Department of Agriculture, Iowa State College, Ames, Iowa, and LEON F. CHARITY is with Iowa State College.

This is *Journal Paper no. J-3257* of the Iowa Agricultural and Home Economics Experiment Station, Ames, Iowa, Project no. *1282*.

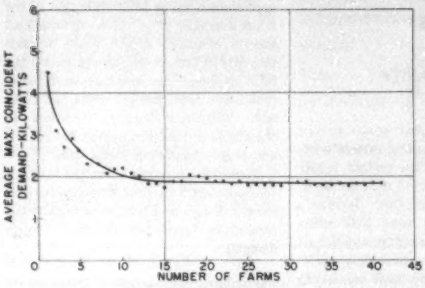


Fig. 1. Variation of average maximum coincident demand per farm with number of farms, Iowa, 1954

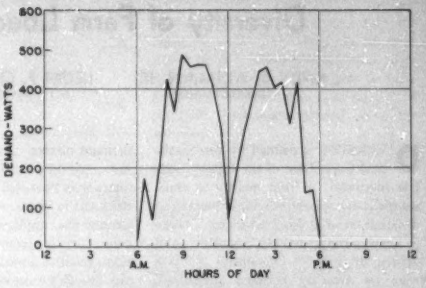


Fig. 3. Average demand of the motors on five elevators, Iowa, October 30–November 5, 1954

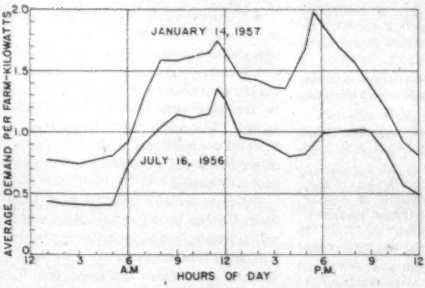


Fig. 2. Daily load curve of generation plant serving rural area in Iowa

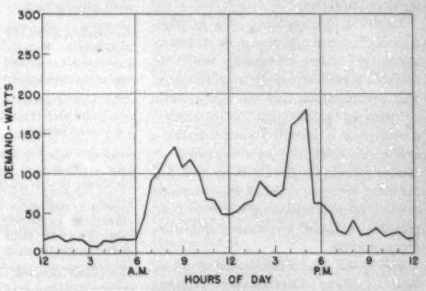


Fig. 4. Average demand of the motors on six water pumps, Iowa, December 1–14, 1954

The equation accounted for 95% of the variation found in the annual maximum demands of these farms. It should be kept in mind that this equation applies only to highly electrified farms with the same type of agriculture and, possibly, geographic location as the farms in the study.

To make this method of transformer sizing more generally applicable, further work is being done in which data from 200 farms is utilized and an electronic computer is employed to solve the equation coefficients. This second group of farms is also located in Iowa.

Diversity Among Farms

The annual diversity factor among the group of 42 farms with ranges was 2.75. The annual diversity factor among the 36 highly electrified farms was 1.98. By definition this means that the sum of the maximum demands of the two groups of individual farms averaged 2.36 times the maximum demands of the farms considered as a group.

There is normally more diversity among a large number of farms than a smaller number. It may be seen from Fig. 1 that the average coincident maximum demand decreased as the number of farms was increased.

Load curves of groups of farms furnish information for diversity studies. In Fig. 2 are daily load curves for July 16, 1956, and January 14, 1957, of a generation plant serving a rural area in Iowa. Promoting the use of appliances that will fill in the low parts of the curve is a problem facing rural distribution system management.

Diversity Among Appliances

Power suppliers are often concerned with the time of the maximum demand of particular types of appliances in relation to the peaking time of the distribution system. Average daily load curves of appliances were prepared to show the demands of appliances. Whether or not the appliances considered in this study smooth out the system daily load curve, i.e. improve load factor, may be determined by a comparison of Fig. 2 with Figs. 3 through 8.

Many of the appliances used in farm production are seasonal in time of operation. To obtain data for the load curves, appliances were metered, in so far as possible, at the season of peak use of the appliance being considered. Since the number of appliances which could be metered at one time was limited, the data for periods of from 1 to 2 weeks were averaged to obtain the daily load curves presented. Smoother and probably more reliable curves resulted.

Grain Elevator Motors

Five grain elevator motors were metered from October 30 to November 5, 1954. Typical use was for 5 to 10 minutes each half hour from 7:30 a.m. to noon and from 12:30 to 5 p.m. It required an average of 7 minutes to unload a load of corn. The average demand in watts for the motors by hours of the day, omitting Sunday, is shown in Fig. 3. Of the motors metered, two were 5 hp, two were 3 hp, and one was 2 hp.

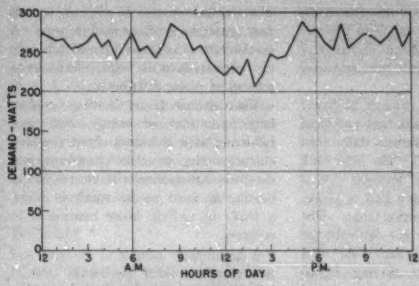


Fig. 5. Average demand of the heating elements on eight stock waterers, Iowa, February 2–15, 1955

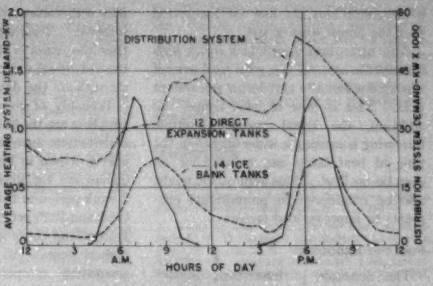


Fig. 7. Daily load curves of bulk milk tanks and a rural distribution system, Iowa, February 1957

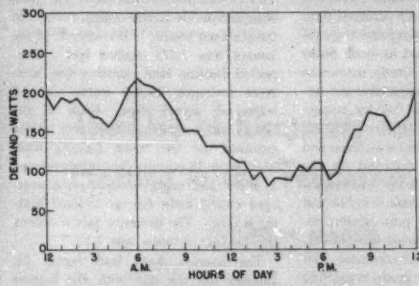


Fig. 6. Average demand of the heating elements on four poultry brooders, Iowa, April 16–26, 1955

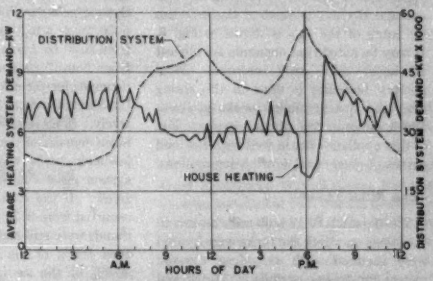


Fig. 8. Daily load curve of eight electric house-heating systems and a rural distribution system, Iowa, January 16, 1957

The daily maximum demands of Iowa rural distribution systems in October and November occur from 6 to 6:30 p.m. This is after the usual time of elevator use. The elevator motors made almost no contribution to the daily, monthly, or annual maximum demands of distribution systems.

Water Pump Motors

The motors on the pumps of six pressure water systems were metered from December 1 to 14, 1954. During these dates the distribution system peaked from 5:30 to 6 p.m. Fig. 4 shows the average demand of the water pump motors by hours of the day, and it should be noted that the group peak demand of the motors did not occur at the time of the system peak.

One of the many advantages of water systems is the resulting improvement in living standards of farm people and income producing possibilities of the farms. Many kinds of electric equipment with excellent load characteristics are associated with pressure water systems, such as water heaters, stock waterers, garden watering, and others. Use of this appliance should be vigorously encouraged.

Heated Stock-Watering Equipment

The use of heating elements for keeping stock-watering equipment free of ice is one of the more common farm applications of electricity in Iowa. In the study of 36 farms, 16 had 25 electrically heated poultry waterers, 10 had 10 electric immersion stock-tank heaters, and 21 had 36 electrically heated automatic stock waterers.

The thermostatically controlled heating elements on seven automatic stock waterers and one stock-tank heater were metered from February 2 to 15, 1955. The 1,000-watt stock-tank heater was installed in a 12-foot-diameter wooden tank. During the 2-week period the heater operated almost continuously. Heating elements on the seven automatic waterers averaged 666 watts in capacity. The elements rarely operated continuously for more than 30 minutes.

Fig. 5 shows the average daily load curve of the eight stock waterer heaters. The almost constant average load with some dropping off during the late morning and afternoon should be noted. These devices have excellent load characteristics and should be considered in promotional programs.

Heat Lamps for Pig Brooding

Pig brooding with infrared lamps is another of the widely used applications of electricity. Thirty farms in the study of 36 farms made use of this appliance in raising hogs. Most heat lamps were operated in February, March, and April. Pigs were farrowed on some of the farms during all seasons of the year, and on these farms heat lamps were used during all of the colder months.

The demands of the heat lamps on four farms were metered for a 2-week period in late March and early April 1955. The lamps on three of the four farms operated continuously during this period. The lamps were turned off during the daylight hours on the other farm. A curve of the pattern of use of this appliance would be

a horizontal straight line for the period that the lamps were in use.

Any utilization item which has a high load factor in itself is a good load even if it is used seasonally and during the system peak. This is true of heat lamps. In addition, the trend toward multiple farrowing is causing a wider spread in the time of heat lamp use, making for an even better load. This application might also be considered in promotional programs to improve load factors.

Poultry Brooders

The demands of hover-type poultry brooders on four farms were metered from April 16 through April 26, 1955. The wattages of the heating elements were 1,300, 1,000, and two of 750 watts. The average electric demand of the brooders by hours of the day is shown in Fig. 6. It may be noted that demands are highest at night and early morning. Most poultry brooding is done in the spring after the system annual peaks in Iowa, which occur in the winter months. Use of this appliance fits in well with the load curves of most rural distribution systems.

Bulk Milk Tanks

The demands of 17 bulk milk tanks and the farms on which the tanks were located were metered with recording demand meters for 2-week periods in March and August 1956. In February 1957 the demands of 9 of the foregoing and 17 additional bulk milk tanks were metered for a similar period of time. The maximum 30-minute demands of the 15 ice-bank units averaged 1.58 kw. The maximum demands of the 19 direct-expansion coolers averaged 2.49 kw. It should be kept in mind that in this section of the United States most of the ice-bank condensing units were of every day pickup size, whereas most of the direct-expansion units were sized for every other day pickup. Had the ice-bank tanks been of the every-other-day pickup size, the condensing units would have been smaller and the maximum demands lower. Sizes averaged 333 gallons for the ice-bank tanks and 341 gallons for those of the direct-expansion type.

March and August data showed that on none of the individual farms did the 30-minute maximum demand of the farm and the cooler coincide. The motors on the ice-bank tanks were running an average of 50.5% of the time during the period of the farm maximum demand. The motors on the direct-expansion tanks averaged being in operation 54.5% of the time during the farm peak. Individual percentages ranged from 0.0 to 90.9 with values scattered throughout the interval. Attention is called to the fact that the times of maximum demands of the individual dairy farms were scattered throughout the day.

Probably of greater interest to power suppliers are the bulk milk tank and rural distribution system average daily load curves shown in Fig. 7. The milk tank curves are the average demands over a 2-week period in February 1957 of 14 ice-bank and 12 direct-expansion tanks. The system load curve is for the day with the highest demand in February 1957 of a generation co-operative serving 29,000 farms.

The group peak demands of both types of bulk milk tanks occurred after the distribution system maximum demand. System peaks in February occurred from 5:30 to 6 p.m., direct-expansion peaks from 6:30 to 7 p.m., and ice-bank peaks from 7 to 7:30 p.m. Group maximum demands for the direct-expansion and ice-bank tanks were 1.28 and 0.76 kw, respectively. The per cent of the group maximum demand of the direct-expansion and ice-bank tanks which occurred on the system peak was 64.6 and 59.9, respectively. If the system peak demand had occurred from 6:30 to 7 p.m. (system demands were still high during this period), then 100% of the direct-expansion and 96.0% of the ice-bank group maximum demand would have occurred on the system peak.

For the ice-bank and direct-expansion tanks, respectively, the diversity factors were 1.97 and 1.96, group load factors 44 and 23%, and group maximum demands at the time of the system peak were 0.45 and 0.83 kw, respectively. This means that a direct-expansion tank of the average size used in this study might be expected to add 0.38 kw more to the system peak than a similar sized ice-bank unit.

A complete report on this study is being published as a research bulletin by the Iowa Agricultural and Home Economics Experiment Station.

House Heating

Several hundred farm homes in Iowa are being heated electrically with most of the heating systems installed within the past 2 years. Resistance cable in the ceiling is the most popular type of construction. The heating installations usually have controls to reduce the voltage to the heaters from 240 to 120 for a period of approximately $1^1/_2$ hours at the time of system peak. Many power suppliers are encouraging this load even though distribution system peak demands occur in winter in Iowa and the off-peak controls may create an artificial system peak. A special rate of between 1.3 and 1.6 cents per kilowatt-hour is being offered by a number of power distributors.

Since electric house heating involves large quantities of energy and since relatively little is known about the load characteristics, structural problems, and consumer satisfaction with electric house heating in areas as far north as Iowa, a study of electric house heating is in progress.

A preliminary sample of 10 farm and two rural nonfarm electrically heated houses was metered with recording demand and kilowatt-hour meters during much of the 1956-57 heating season. Ten of the houses were new and used resistance cable in the ceiling. The old houses used panels. Average size of the houses was 1,373 square feet. Connected heating load omitting the basement averaged 14,829 watts, or 10.8 watts per square foot. Each of the houses had a separate 100-ampere service entrance for the house heating load. Maximum 15-minute demands averaged 11.8 kw and energy consumption averaged 16,303 kwhr for the 1956-57 heating season. The diversity factor among the heating systems was 1.23.

The average daily load curves for January 16, the day with the highest house heating coincident maximum demand, of the heating systems of eight of the houses having off-peak controls and a distribution system serving a rural area are shown in Fig. 8. It may be seen that the off-peak controls are effective in reducing the demand at the time of the system peak.

The house-heating peak demand immediately following the control period may be noted. From a comparison of data from this limited study and the distribution system daily load curve shown in Fig. 8, it is estimated that an artificial system peak would be created after the control period if house-heating saturations are above 4.8%. Different saturations would be required to create artificial peaks if the shape of the daily load curve of the system is different from the one used in this study.

The annual load factor of the 12 electric house-heating systems in this study averaged 15.9%, a little less than typical well-equipped farms. Values ranged from 12.1 to 19.4%. Group annual load factor of eight of the heating systems with off-peak controls was 18.9%. Off-peak controls tended to lower the annual group load factor by causing much of the connected load to be in operation after the

control period. If off-peak controls had not been used, computations indicate that the group annual load factor would have been about 22%. If the demand of the group at the time of the system peak were assumed to be the maximum demand, the annual load factor would then be 50.3%.

The owners of the electrically heated houses seemed well pleased with the operation of their heating systems from both a cost and comfort viewpoint.

Summary

Data obtained by metering the demands of farms, homes, and appliances have been presented in this paper. The following statements are made to emphasize certain points on the diversity of farm loads:

1. The diversity factor between the maximum demands of equipment in the farm home and in the other buildings was approximately 1.4. On farms considered as a group there was little diversity between house and production appliances at the time of the distribution system peak.

2. Elevator and crop dryer motors and poultry brooders had little possibility of adding to the annual distribution system maximum demand. Equipment such as stock water heaters and heat lamps for pig brooding had excellent load factors, even though they may have added to the distribution system peak.

3. Ice-bank and direct-expansion bulk milk tanks had diversity factors of 1.97 and 1.96 and group load factors 44 and 23% respectively. On the average, each direct-expansion bulk milk tank on the distribution system added 0.38 kw more to the system maximum demand than a similar sized ice-bank unit. Condensing units for ice-bank tanks were mostly sized for every day milk pickup; direct-expansion tanks were sized for every other day pickup.

4. There was very little diversity among the electric house heating installations. With present system loads and as long as house-heating saturations are low, electric heating with off-peak controls will improve system load factors. As saturations increase, controls will cause an artificial system peak.

References

1. Glossary of Electric Terms—Financial and Technical. Statistical Committee, Edison Electric Institute, New York, N. Y., 1956.

2. Demand and Diversity of Use of Electricity on 16 Farms in the Eastern Livestock Area of Iowa, L. B. Altman, Kathryn Philson, E. J. Bureh. *Research Bulletin 387*, Iowa Agricultural Experiment Station, Ames, Iowa, May 1952.

3. Load Characteristics of Southeastern Iowa Farms Using Electric Ranges, L. B. Altman, E. H. Jebe. *Research Bulletin 420, Ibid.*, Jan. 1955.

4. Glossary of Important Power and Rate Terms, Abbreviations, and Units of Measure. Federal Power Commission, et al., U. S. Government Printing Office, Washington, D. C., 1949.

The Use of Electricity for House Heating in the Tennessee Valley

BUFORD H. MARTIN
MEMBER AIEE

TODAY several hundred thousand homes throughout the country are being heated satisfactorily with electricity and the number is growing quite rapidly. Five years ago the majority of electrically heated homes were concentrated in only a few areas, but now the use of electric heating is spreading into many places where there was none as recent as 2 years ago. A number of factors are responsible for this widening use. Among the more influential of these is the increasing cost of fuels and the tremendous growth of summer air-conditioning loads on electric distribution systems. Many electric utilities are actively promoting heating to offset the summer peaks created by air conditioning.

In the Tennessee Valley, the number of homes using electric heating is rapidly approaching 200,000. The most recent estimate indicates over 177,000 homes with complete electric heating systems. This represents a saturation of about 14%. It is not known how many homes use electricity for auxiliary heating, or for "chill-chasing" in the fall before the fuel-fired furnace is placed into operation, or in the late spring after the furnace has been shut down.

Fig. 1 shows the annual growth in the number of saturation of electrically heated homes in the area served by the distributors of TVA (Tennessee Valley Authority) power from 1948 to the beginning of 1957. Fifty-one of these electric systems are co-operatives.

As might be expected, the majority of these electrically heated homes had the systems installed at the time the home was built. Generally speaking, the old houses are poorly insulated, and conversion to electric heating is expensive; however, electric heating has been installed in older houses when they were well constructed and after suitable thermal insulation was provided.

Some indication of the number of new electrically heated homes built during the last 5 years in areas served by the co-operatives in the Tennessee Valley may be obtained from Table I, which shows the annual increase over the previous year in residential consumers and electric heating installations in TVA areas served by co-operatives.

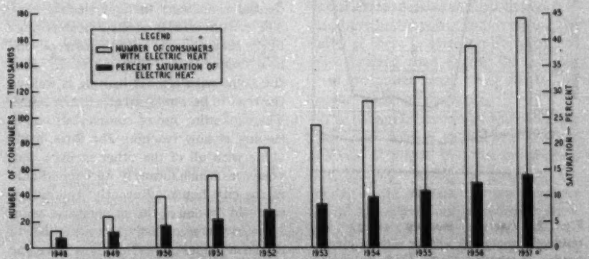


Fig. 1. Saturation of electric space heating among residential consumers served by municipal and co-operative distributors of TVA power

Paper 57-1161, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Farm Electrification Conference, Minneapolis, Minn., October 29–31, 1957. Manuscript submitted September 6, 1957; made available for printing May 26, 1958.

Buford H. Martin is with the Tennessee Valley Authority, Chattanooga, Tenn.

Table I. Growth of Residential Consumers and Electric Heating Installations

January	Residential Increase	Electric Heat
1953	22,400	4,700
1954	16,200	3,900
1955	13,900	3,700
1956	9,300	5,200
1957	11,270	5,500

It may be noted that in the last 2 years the increase in electrically heated homes was about 50% of the new consumers added during the same years. In some of the municipal areas, 80% to over 90% of the new homes built within the last 5 years are heated electrically.

A utilization specialist working with one of the rural electric co-operatives in northern Alabama reported that of 105 homes for which he made electric space-heating recommendations during 1956, 103 installed complete electric heating. This co-operative, whose members are over 90% rural, is located in an area where bottled gas competition is very strong. Its average annual residential kilowatt-hour use increased from 1,384 in 1950 to 4,502 as of June 30, 1957. This 225% increase in average annual use has resulted largely from increased use of electricity for house heating along with aggressive promotion of major home appliances.

This increase has occurred even though other means of heating were considered to be cheaper, which is generally true from a cost-of-operation standpoint, but there are many offsetting advantages to electric heating which are greatly appreciated by the users. Among the more important of these are cleanliness, accurate temperature control, and the almost complete freedom from attention during operation. The cleanliness of electric heat has a strong appeal to the housewife. The walls of an electrically heated home do not require redecoration as frequently as when other heating systems are used. Draperies, curtains, and rugs do not require cleaning and laundering nearly so often. All of this represents tangible savings and should be considered as a credit when estimating the over-all cost of electric heating in the home.

Also, there is the advantage gained from the heat supplied by lighting, television, radio, and other electric appliances. Since these basic loads often supply a substantial portion of the heat requirements during mild weather, the electricity required for the heaters can be purchased at lower steps of the rate. In an electrically heated home there is no excuse for the family's not having the best of lighting. When the current is used to produce light, heat is also obtained, and therefore less current will be used by the heaters. With the increasing use of better lighting and appliances in the home, electric heating is made even more attractive.

Surveys, made in co-operation with some of the TVA power distributors, indicate very definitely that the consumers having electric heating are better year-round users of electricity. This is well illustrated by Fig. 2, which shows the average monthly energy use by residential consumers with or without electric heating as reported by 13 TVA distributors. From this it is seen that the average summer use by the residential consumers having electric heat is over twice that of the residential consumers without electric heat. The consumer with electric heat is more likely than the one without to cook and heat water with electricity. In fact, the home with electric heating is well on the road to becoming an all-electric home. The expanding use of summer air conditioning is now reaching the farm home along with all of the other modern conveniences which formerly were found only in the city home. Recently, the author was told by one of the co-operative managers in the area that quite a number of the farm homes on his lines now have one or more room coolers, and some have installed complete air-conditioning systems. As this summer load grows, it will tend to produce a better year-round load in the home.

Most important in making electric heating feasible to both the user and the power supplier is proper application of a well-selected system for the type of construction of the home in which it is to be installed.

Since electricity is a high-cost form of energy for space heating, good building insulation is imperative. To install electric heating in poorly constructed or inadequately insulated homes will almost certainly result in dissatisfaction to the user and will be expensive to the power supplier. Fig. 3 shows the effect of insulation on the capacity required for heating. This does not, however, reflect the added comfort which may be realized from good insulation.

Electric space-heating systems may be generally classified into two groups: central systems and those consisting of heating units installed in each space to be heated. With the central system, the heat is transferred to the various rooms by means of circulating warm air, water, or steam, and usually the regulation is by only one thermostat located at some central place in the house. With the use of separate units in each space, the temperature of each room can be, and usually is, individually controlled by wall thermostats located in the room or built into the heater itself. Because of the independent control for each space, systems using unit heaters will generally use less electricity

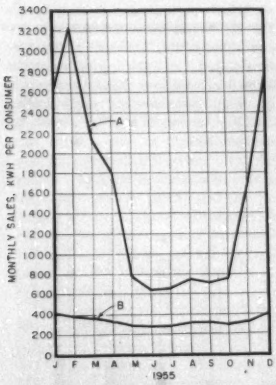


Fig. 2. Average monthly energy use, residential consumers with or without electric heating as reported by 13 TVA distributors

A—With electric heating
B—Without electric heating

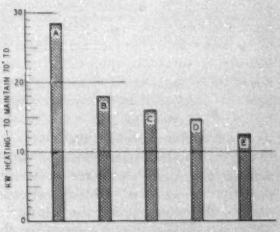


Fig. 3. The effect of insulation on heating capacity required in 3-bedroom residence with 8,200 cubic feet of heated volume. Frame construction:

A—With no insulation
B—With 3½-inch mineral wool over ceiling only
C—With 3½-inch mineral wool over ceiling and 2-inch insulation in walls
D—With 3½-inch insulation in walls and over ceiling, plus weather-stripping of doors and windows
E—Same as D with addition of double glass in windows

and have better load characteristics than central systems. Better regulated temperatures throughout the house can be had with individual room control. Heat will be provided in each room as needed. The shady side of the house or that part of the building toward a cold wind will require more heat than the rooms exposed to the sun. Full advantage of the heat radiated from the sun cannot be utilized with a central heating system.

There are several kinds of unit heaters. These may be generally classified into six groups: wall insert heaters, baseboard heaters, portable heaters, radiant panels, floor furnaces, and resistance wire embedded in the ceiling or floor. There are forced circulation heaters, radiators, radiant-convection heaters, and strictly convection types. There are good and better heaters of each type and the one to choose for any particular job may depend on such factors as the type of building construction, local availability, personal likes or dislikes as to style of the heaters, and the price. A combination of type and make of heaters is often used.

The growing demand for summer cooling will influence the extent to which central systems are installed, because they provide a means of filtering or electrostatic cleaning of the air in addition to controlling temperature. The heat pump for year-round air conditioning will certainly become more generally used in the larger and more expensive homes. Any modern new home, however, for which air conditioning is contemplated is a logical prospect for a heat pump installation, no matter what its size.

The heat pump is now an approved year-round air-conditioning device well beyond the experimental stage. Several reputable manufacturers are now making heat pumps and soon it may be expected that others will be added to the growing list. The number of heat pump units of all types now operating in the Tennessee Valley is estimated to be near 2,000. Approximately half of these are in home installations. A large number of the units are on the lines of the co-operatives. One installation in a chicken hatchery has received wide publicity, and articles dealing with this application have been published. Eighteen of the municipal distributors and 16 of the co-operatives distributing TVA power are using heat pumps in one or more of their own buildings. In the Tennessee Valley, 34 power distributors now have heat pump installations in their offices.

Technical data for use in making the computations for the capacity of the heating system to be installed are available from several sources. The *Heating, Ventilating, Air Conditioning Guide*[1] contains tables giving heat loss coefficients of many building and insulating materials. The chapter pertaining to electric space heating for the 1958 issue of the Guide is now in the process of revision.

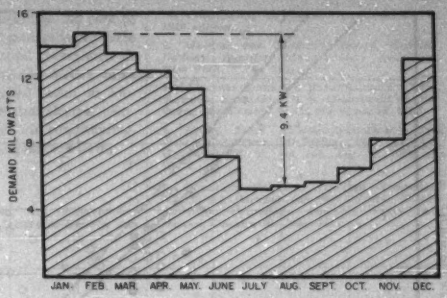


Fig. 4. Composite average monthly demands of five electrically heated homes on a rural co-operative system in Tennessee. Average heating load: 20 kw

The Electric House Heating Equipment Section of The National Electrical Manufacturers Association has available the *Manual for Electric House Heating*,[2] which is intended as a basic guide for the design of electric heating installations. It also contains a chapter which discusses the method of estimating annual kilowatt-hour consumption. Several manufacturers of space-heating equipment have issued instruction booklets for use in determining building heat loss and designing heating systems for the proper application of their products.

The heat loss is influenced by a number of variables, such as geographic location, exposure to winds, design of the building glass areas, number of outside doors, and allowance for air infiltration. The amount of electricity used is also subject to such influences as the temperature maintained inside the house, frequency of opening outside doors, number of people living in the home, wind velocities, hours of sunshine, and several others.

A commonly used equation for estimating the annual kilowatt-hour consumption for any specific installation is

$$\text{annual kilowatt-hours} = \frac{HL \times DD \times C}{TD}$$

where

HL = calculated heat loss of the house, kilowatts

DD = annual degree days for locality in which the home is located

C = constant representing average hours per day during heating season for which the full heating load may be expected to operate. It can never be over 24 and usually varies between 15 and 20

TD = difference between inside and outside design temperatures in degrees Fahrenheit

Having determined the annual kilowatt-hour consumption by this method, the annual cost may be estimated by applying the local power rate.

The number of degree days for a normal heating season in Knoxville, as given in the *Weather Bureau's Climatological Data*,[3] is 3,590. Assuming a design temperature difference of 70 degrees and an average operating time per day of 18.5 hours for the full heating season as the value of C, a house having a heat loss equivalent to 15 kw (50,000 Btu-hours) would require 14,232 kwhr (kilowatt-hours) to heat it during a normal winter. A central system having only one thermostat control might require from 15% to 20% more. The demand characteristics of space-heating loads are of particular interest to the power supplier. He is concerned as to how these affect both his investment required to render service and the annual load factor of his system.

Electric resistance heaters have, for all practical purposes, unity power factor. Resistance space-heating loads are therefore beneficial in helping to improve system power factor. It has been observed that, in general, distribution systems have better power factors during the winter if there is considerable resistance space heating than during the summer when low power factor room coolers and similar loads predominate.

The demands created by space-heating loads are highest during the coldest weather, at which time the distribution facilities are capable of handling the greater loads. Where a group of houses can be served from one transformer, there is the advantage of considerable diversity of the heating and appliance loads. On rural lines where only one farm is served from one transformer, the diversity factor may still be surprisingly good if the farm

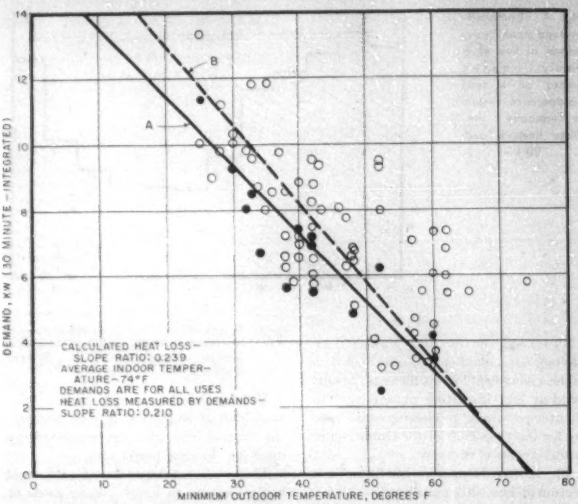


Fig. 5. Demands measured with Westinghouse watt-hour meter, RW indicating register with 30-minute time interval

A—Indicated heat less from demand data
B—Calculated heat loss at 65-degree temperature difference
●—Reading taken during night period: heating, refrigerator, water heater, and food freezer only
○—Reading taken during day and night with other loads, lighting and cooking in use

is well equipped electrically. Often the transformer capacity required to serve farm machinery, such as hay dryers, feed grinders, and similar equipment, will be adequate to serve the space heating load in the house. The heating load creates its heaviest demand at times when the electric machinery is not operated. One cooperative manager has established a formula to govern the transformer capacity required to serve a farm home having electric space-heating. This is one half the installed heating capacity plus 4.8 kw for other household uses. A transformer rated nearest to this figure is selected. He has been using this rule successfully for several years.

Fig. 4 shows the composite average monthly demands of five electrically heated homes on this co-operative manager's system. The average connected space heating load of the five homes is 20 kw. Applying his formula, the calculated size of the transformer would be a 15 kva. Note that the highest demand shown on the chart is 14.8 kw which occurred during February. Note also that this is only 9.4 kw higher than the demand during the summer months. All of the houses are insulated with mineral wool: 4 inches over the ceiling and in outside walls. Four houses are of brick-veneer construction an one is frame-type with siding. Two of the houses have 6 rooms, two have 7, and one has 9. The highest demand for this 9-room house was 19 kw with a summer demand of 5 kw. The connected space heating load is 28 kw. All heating installations are of the individual room unit type.

The average annual kilowatt-hour consumption for the group was 19,700, and the estimated amount required for heating was heating was 15,100 kwhr. The normal heating season for this area has 3,500 degree-days.

The average annual load factor of the group was 15.3%, but without the heating and based on the summer demand, it was only 9.8%. The annual load factor of the resistance heating alone was 18.3%.

Fig. 5 shows demand readings as related to minimum outdoor temperatures taken in December 1956 and January 1957 in a home having a computed heat loss of 15.5 kw at a 65-degree temperature difference and an installed capacity of 16 kw. A special metering arrangement was installed and readings were taken at various intervals, at which time the demand indicator would be reset. A heat loss line having a slope ratio of 0.210 was determined by readings made when the family retired and all lights, television, radio, and cooking applicances were off, then again early in the morning before any of these appliances were turned on. Temperature readings both inside and outside of the house were made at the same time. It may be seen that the slope of the actual heat loss follows rather closely that made by calculation based on heat loss formulas. Note also that the highest demand taken during the period was only 7.5 kw over that established in warmer periods by the basic loads in the home. This difference is 47% of the 16-kw installed heating load. The minimum outdoor temperature during the period when the highest demand was registered was 23 degrees. The temperature indoors during this period was 74 degrees, a difference of 51 degrees or 78% of the design difference. The calculated increase in heating demand for an outdoor temperature of 10 degrees is 1.7 kw, which would raise the total demand to 16 kw.

Now, what effect is space heating load having upon the annual system load factor of co-operatives as the saturation of house heating increases? Table II shows the annual load factor for 1956 of five co-operative systems in the TVA area having the highest saturation of space-heating load compared with the five co-operative systems having the lowest saturation of house-heating load.

These figures indicate that house-heating saturations up to 15% do not significantly affect the system's annual load factor.

Indications are that the future growth of electric house heating will be widespread and greatly accelerated. Certainly it should be expected in this area. The emphasis on the comfort, safety,

Table II. Annual Load Factor of Systems With High and Low Saturation of Electrically Heated Homes

System	Number of Heating Consumers	Saturation, Per Cent	Annual Load Factor, Per Cent
A	3,100	18.4	49
B	2,500	14.5	57
C	2,900	14.4	49
D	2,300	14.4	52
E	600	11.7	52
Average		14.7	51.8
F	100	1.0	52
G	80	1.0	54
H	50	1.0	51
I	40	1.1	50
J	20	1.0	55
Average		1.0	52.4

cleanliness, space saving, and carefree operation which result from the use of electric home heating, is causing the public to demand it. A housewife living in a farm home served by one of the cooperatives in Alabama stated that before using electric heat she had had to "fuss with a coal stove in each room" of a large farm house, which increased her housekeeping chores considerably. Since they have moved into their new electrically heated home she is "enjoying a clean house with no fires to kindle, no coal to carry in, and no soot and ashes to mess up the place." Satisfied users such as this one in the Tennessee Valley have been the most influential factor in convincing home builders that electric heating, and (incidentally) air conditioning, are essential in providing modern living.

References

1. Heating, Ventilating, Air Conditioning Guide (book), American Society of Heating and Air Conditioning Engineers, New York, N. Y., 1956 edition.

2. Manual for Electric House Heating. National Electrical Manufacturers Association, New York, N. Y., May 1956.

3. Weather Bureau's Climatological Data, National Summary. U. S. Department of Commerce, Washington, D. C., vol. 6, no. 13, 1955.

4. Heat Pump Goes Industrial; Cuts Hatchery's Costs, J. D. Brown. *Electrical World*, New York, N. Y., vol. 145, Apr. 30, 1956, pp. 70-71.

5. Electric Heating in the Tennessee Valley—A Pattern for the Future, William R. New. *Proceedings*, American Power Conference, Chicago, Ill., Mar. 1957.

An Analysis of Visual Elements in Roadway Lighting

W. H. EDMAN
NONMEMBER AIEE

VISUAL elements which are fundamental to roadway* lighting need to be clarified. The many varying elements of night visibility form a complex interrelationship which is not generally understood. Further studies and investigations are being carried on to obtain more quantitative data on such visual elements as discomfort glare, veiling glare, and pavement and object reflectance characteristics. Also, practical methods and means for measurements are currently being developed. However, until all this becomes a reality, we must utilize established facts and proven experience.

Unfortunately, known facts and well-established principles have often become obscured by promotional statements purported to be scientific and factual. This has led many well meaning but uninformed people to adopt changes in practice which have proven impractical and factually incorrect. Such unfortunate mistakes could perhaps have been avoided if the known fundamental principles and factual knowledge had been more clearly set forth by the roadway lighting industry.

This paper sets forth the following important visual elements in roadway lighting:

1. Illumination intensity,
2. Background and object brightness,
3. Veiling and discomfort glare,
4. Uniformity of illumination and brightness,

and presents an analysis of their relation to the fundamental principles of roadway lighting practice.

* This term has been recently adopted by the Illuminating Engineering Society to apply to street and highway lighting.

Requirements

The underlying requirements of roadway lighting can be stated quite simply. The first and most important function of roadway lighting is to reveal objects on the roadway proper and sidewalks. The driver's field of view is small in angular size as illustrated in Fig. 1. It consists of the roadway, sidewalks, buildings, and other vertical surfaces such as trees, fences, etc., the sky, luminaires, and the atmosphere when hazy. All these surfaces form a picture of changing and varying brightnesses. Then, we must introduce into this scene the objects which are to be found in action, particularly on the roadway. The motorist and pedestrian must evaluate the whole situation and the interrelation of each element, one to another, and the total effect. Mental estimates must be made as a basis for judgement as to relative speeds, distances, object size, curvature, and solve subconsciously the complicated relationship between perspective and plan. The average traffic condition presents many objects of different shapes and dimensions. Their brightnesses shift and vary according to a complex distribution pattern. An infinite number of changing visual conditions are constantly being presented. The observer then must interpret this entire complex scene and make split-second decisions to initiate subsequent actions necessary for safe maneuvering.

The effectiveness of the lighting system will be the extent to which it fulfills the basic function of safely revealing without discomfort. When the economics of street lighting is a serious matter, then the creation of maximum visibility distance is of primary importance and comfort is quite secondary. While comfort and pleasing appearance are desirable, they should never be obtained at the expense of necessary visibility distance.

In dealing with the safety aspects of roadway lighting, the requirements, problems, and questions are quite clear. However, many of the answers are complex. We know who uses the streets and highways. We also know what the users are looking for, where they are, and what needs to be seen. These needs are quite simple and can be considered step by step.

When the necessary visibility distance for modern traffic conditions has been provided, then we can assume that all other safety requirements of roadway lighting have been satisfied. A motorist must be able to see where he is going and what is going on within his immediate area. Where the road goes, where the road stops, where the road is clear, are basic to safety. Any object must be clearly seen and its behavior ascertained at all times. Important things must be seen first.

Visibility Defined

The concept of visibility as used in roadway lighting has never been clearly defined simply because it cannot be given a precise meaning. The term is loosely

Paper 58-549, recommended by the AIEE Production and Application of Light Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Middle Eastern District Meeting, Washington, D. C., April 28-30, 1958. Manuscript submitted January 21, 1958; made available for printing April 8, 1958.

W. H. Edman is with the Holophane Company, Inc., Newark, Ohio.

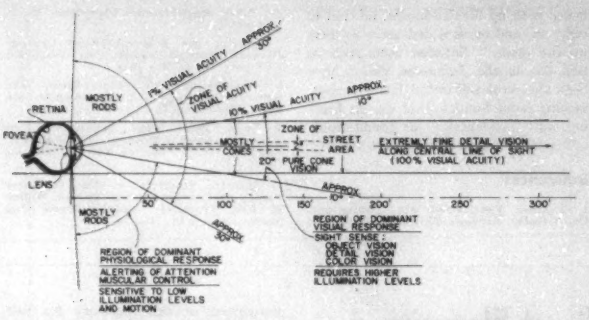


Fig. 1. Relationship of zones of vision of the human eye to a typical street area

used in many different senses. It is somewhat like trying precisely to define the term love. Visibility is such a broad term that it must be broken down into component factors which can be given specific meaning. P. S. Millar,[1] to the author's knowledge, was the first to use the term "revealing qualities" of street illumination. Later, J. M. Waldram[2] of England, used a similar approach in trying to give visibility a more clearly understood meaning and proposed a method for its evaluation. It might also be thought of as the capacity of a lighting system to produce visibility distance under given conditions. Fig. 2 is a modification of Waldram's original chart showing the relationship of many elements which must be dealt with and given consideration in the problem of roadway lighting. A mere glance indicates at least 25 elements which influence visibility distance.

The Human Eye

In general, seeing conditions are largely a matter of controlling the distribution of brightnesses in the field of view. Actually, in street lighting, the field of view is generally over long distances. The central zone of the eyes of a motorist is constantly scanning from one side of the roadway area to the other, from the immediate foreground to far ahead and back, to guide him safely along his path of motion. We are not generally concerned with the ultimate possible detailed vision. Such "concentration" would be too hazardous. Vision on streets in daytime or nighttime is usually of the order of medium detail in about a 20-degree-wide region as shown in Fig. 1, with occasional finer detail seeing required on nearby objects, signs, etc. For this type of vision a generally higher level of illumination is needed since this is predominantly in the zone of cone vision.

Accordingly, there should be sufficiently high levels of illumination and brightnesses in the street area where most of the activity is taking place and where detail or cone vision will be concentrated. In fact, if there is not sufficient illumination on objects and immediate surround in the street area, in contrast to illumination in the peripheral regions, it will be difficult for the eyes to concentrate vision within the street area. They will be distracted.

On the other hand, cone vision alone would be dangerous. The eye is equipped with its very sensitive peripheral or rod vision to give the alerting of attention which allows man to survive in his environment. These rods in the peripheral field are very sensitive in detecting motion at low levels of brightness (at higher brightness levels the cone receptors dominate). For example, under conditions of bright moonlight [0.01 foot-lambert (ft-l)] the effective visibility on an open highway is considerably enhanced as compared to a dark cloudy night.

The characteristics of rod and cone vision establish the criterion that the maximum obtainable percentage of the luminaire output should be directed onto the roadway area. This will provide greater visibility distance and prevent distraction from the surround. For proper balance the surround illumination requirements will always be fulfilled by the stray light from the luminaires and reflected light from the roadway, sidewalks, and adjacent surfaces.

Methods of Sight on Roadways at Night

The method of discernment of objects in the field of view are generally divided into two types. When the background brightness is greater than the brightness of the object to be seen, it is known as silhouette discernment. When the opposite condition exists, we call it reverse silhouette discernment. The later is also referred to as discernment by surface detail, or illumination effect. It should be noted that all objects on the roadway are seen because of the brightness contrast of the object surface or

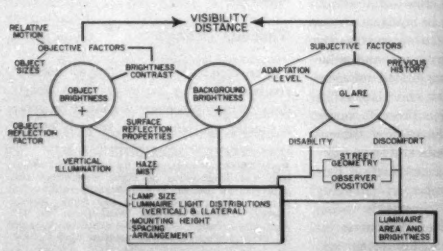


Fig. 2. Interrelationship of elements effecting night visibility distance

Fig. 3. Night view of a traffic street

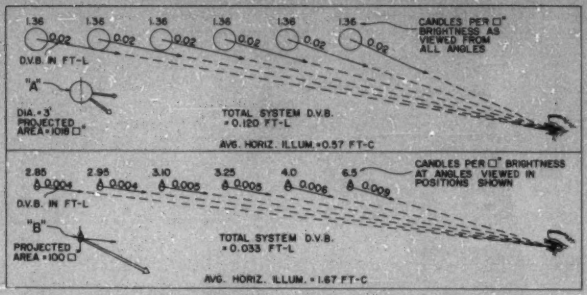


Fig. 4. Disability veiling glare comparison—lamp size 13,000 lumens, street width 50 feet, luminaire spacing 100 feet, mounting height 30 feet, overhang 5 feet, observer eye level above pavement 4½ feet, viewing position 150 feet behind first observed luminaire (typical 76-degree windshield cutoff obscures the luminaire 50 feet ahead of observer). Note the controlled light distribution from the smaller luminaire "B" produces lower system disability veiling glare and higher pavement illumination than does the larger luminaire "A" with uncontrolled distribution

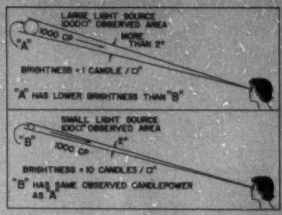


Fig. 5. Discomfort glare at short distances. Larger light source "A" has lower discomfort glare effect than smaller light source "B" because brightness per square inch is lower (all other factors remaining equal)

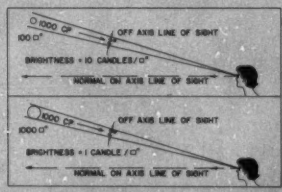

1000 CP
100 □"
OFF AXIS LINE OF SIGHT
BRIGHTNESS = 10 CANDLES / □"
NORMAL OR AXIS LINE OF SIGHT
1000 CP
1000 □"
OFF AXIS LINE OF SIGHT
BRIGHTNESS = 1 CANDLE / □"
NORMAL OR AXIS LINE OF SIGHT

Fig. 6. Discomfort glare at long distances. When projected area of a light source lies within approximately 2-degree visual angle (off-axis) discomfort glare then becomes proportional to observed candlepower regardless of source area

parts of the surface with its background. Seldom do we see entirely by either one of these two methods, but rather, by a combination of both.

Silhouette Effect

This method is most effective where high levels of illumination on objects cannot be provided for reasons of economy, such as on very light traffic or residential streets. In such cases, there are generally few vehicles on the roadway, speeds are relatively slow and the open roadway can be seen as a background for long distances ahead. Silhouette discernment can also be important in high level illumination [one foot-candle (ft-c) and above] when objects are viewed at such distances away that it is impossible to have detail discernment upon the object. Under such circumstances, the lighting system should be designed to provide relatively uniform brightness (average to minimum foot-lambert ratio of less than 4 to 1) as well as good horizontal foot-candle uniformity (average to minimum foot-candle ratio less than 3 to 1).

Even though the silhouette effect may predominate on a particular roadway area, the reverse silhouette effect also produced by the street luminaries helps to produce better perspective and greater visibility distance. It aids in producing vision of the entire object. A silhouette condition alone often outlines only the lower half, the top half being lost above the horizon, unless there are sufficient instances where the top and the side surfaces are seen in reverse silhouette to indicate size, shape, and direction of motion.

Reverse Silhouette Effect

In reverse silhouette the objects are seen directly because the reflected light or surface brightness produces sufficient contrast with the darker background. As the general illumination level is increased, the details of the object also become more readily apparent. In general, on business streets and heavy traffic thoroughfares, discernment by reverse silhouette predominates. This is particularly true during rush hours with high density traffic, both vehicular and pedestrian. Relatively high vertical illumination falling on the object is desirable.

The glint aspect of reverse silhouette can be helpful in providing clues, even under very low illumination levels where silhouette seeing predominates. It is produced by reflection of a light source from a specular surface, such as a pool of water, a polished fender, a chrome plated bumper, or the bright button on a dog's collar. It is most effective in alerting the attention of the motorist under a condition of motion. Here, the observer may not necessarily see the entire object but perceives its presence by the glint produced.

There is another aspect to consider; namely, the shadow effect. Often, it is helpful when the observer cannot readily see the object but is alerted to its presence by the shadow it casts.

The photograph in Fig. 3 illustrates the different methods by which we see at night. Incidentally, most night photographs are not realistic because only empty streets are shown. Here is an example of practical difficulties encountered because of traffic and headlight interference. Notice that a part of the police force must be enlisted to take a simple but expensive photograph. In viewing this photograph, note that the pedestrians and vehicles are quite clearly revealed. The lower part of the officer's body in the center of the intersection is revealed by silhouette and most of the upper portion of the body by reverse silhouette. The man in the light colored suit is predominately revealed by silhouette, although not entirely. Attention is also called to the shadow aspect. Glint or specular reflection can be easily picked up from small areas of the parked vehicles. Yes, even the officer's belt buckle can be seen. All of these elements provide clues to aid safe travel.

Glare Effects

The term "glare" is loosely used to designate two entirely distinct and independent phenomena; namely:

1. The immediate ocular discomfort experienced when looking at an intensely bright light and
2. The veiling effect of the light source.

Table I. The Effect of Variable Luminaire Factors on Visual Elements in Roadway Lighting

Variable Luminaire Factors	Visual Elements											
	Foot-Candle Intensity		Foot-Lambert Brightness					Uniformity			Glare	
			Pavement				Object	Ft-c		Ft-l		
			Concrete		Asphalt							
	Horiz. Plane	Vert. Plane	Dry	Wet	Dry	Wet		Horiz. Plane	Vert. Plane	Bright-ness	Veil-ing	Dis-comfort
Area and brightness (constant output)	3	3	2	1	2	1	2	3	3	2	3	1
Candlepower above 60 degrees vertical (within 30 degrees lateral to curb)	2	1	1	1	1	1	1	1	2	1	1	1
Candlepower below 60 degrees vertical	1	1	1	2	1	2	1	1	2	2	2	2
Mounting height	1	1	1	1	1	1	1	1	1	1	1	1
Spacing	1	1	1	1	1	1	1	1	1	1	1	1
Arrangement (staggered opposite one side)	2	2	2	2	2	2	2	1	1	1	2	2
Location (transverse to street)	2	2	2	1	1	1	2	2	2	1	1	1

1—Major effect 2—Minor effect 3—No effect

This latter is also referred to as blinding glare or disability glare. To each, the human eye and body reacts in a different way. While both are dependent on candlepower in the direction of the observer, there is little or no known relationship between the two as far as the eye is concerned.

Veiling Glare

This form of glare is the most important in roadway lighting because it produces the greatest waste in reducing visibility distance. It is dependent on the total amount of flux entering the eye and the angular displacement of the glare source from the normal line of sight as shown in Fig. 4.

The effect of veiling glare is to interfere with the seeing ability of the observer. The vertical foot-candles at the eye are additive from all the light sources, which cause a veiling effect within the eye. This interference reduces the recognition or visibility distance. The physical size of the glare source has no bearing on veiling glare, since it is a function of candlepower and viewing angle.

Discomfort Glare; Figs. 5 and 6

This form of glare is related to the contrast between the source or luminaire brightness and the surround brightness. It is also related to the angular displacement of the source from the observer's normal line of sight. The chief factors influencing discomfort glare are:

1. Brightness of the luminaire.
2. Visual size of the luminaire.
3. Its relative position with respect to line of sight.
4. Background brightness or adaptation level.
5. Time of exposure.

The major effect of discomfort glare is to produce a sensation of ocular discomfort. It has no measured effect on visibility distance.

When surround brightness is increased, the discomfort effect is reduced. If the projected area of the luminaire is increased, see Fig. 6, and the candlepower in the viewing direction remains the same, the brightness value (candles per square inch or foot-lamberts) will be reduced, which, in turn, reduces the discomfort effect. Merely increasing the source size will not improve the ability to see objects at greater distances.

Actually, the beneficial effect of increasing luminaire area applies only when viewing one or two luminaires in the immediate foreground. For example, in a system of luminaires even a 6-foot-diameter luminaire located 175 feet away from the observer has a solid angle of only 2 degrees. When the luminaire has a visual angle greater than 2 degrees, the size of the luminaire and its unit brightness have an effect on discomfort glare. As the luminaire area is increased, the surface brightness is reduced in direct proportion to the increase in area for a given lumen output. However, the capacity to reduce ocular discomfort is not directly proportional to the increase in area. The recent investigations of Putnam and Bower[3] show that the discomfort glare effect was reduced to one third when the area was increased by a factor of 6.35.

Glare Evaluation

It is practically impossible to accurately evaluate either the discomfort or veiling glare effect of a roadway lighting system by merely observing one sample luminaire at close range.

The veiling glare effect from one luminaire or even a system of luminaires can be calculated and evaluated from known data. However, the eye cannot in any way judge by casual observation the amount of loss in visibility distance due to veiling glare. The threshhold visibility distance could be determined by using targets or objects and making comparative observations with luminaires, both exposed and shielded from the eye. The difference in visibility distance between a shielded and unshielded condition is caused by the disability veiling glare effect.

The discomfort glare effect can be observed to some degree from one sample luminaire. However, it is not possible from this to estimate the total discomfort effect of a roadway lighting system. Care must be used not to be misled when comparing a large luminaire with a relatively small luminaire at close range. As previously shown in Figs. 5 and 6, a large area luminaire with equal lumen output would obviously be more comfortable.

Too often, judgement is passed on the effectiveness of a street lighting system as a result of casual visual appraisals. Actually, in such instances, an appraisal is being made only of the discomfort effect at close range and this can be quite misleading when observing only a single luminaire. It is therefore important to recognize the distinction between veiling glare and discomfort glare when making visual appraisals.

Luminaire Factors and Visual Elements

The luminaire factors (listed in Table I) can be controlled by the lighting design and application engineers. Only by varying these factors can the visual elements in roadway lighting be modified in order to produce maximum visibility distance. Lighting engineers, obviously, have no direct control over the physical makeup of the roadway and its surround.

This table is an attempt to show only by arbitrary ratings the relative effect which each luminaire factor can, by variation, produce on the four important visual elements; namely, foot-candle intensity, brightness, uniformity, and glare. No way has been found to measure by

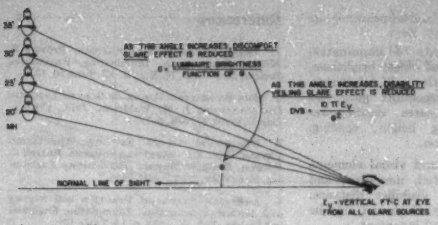

AS THIS ANGLE INCREASES, DISCOMFORT GLARE EFFECT IS REDUCED
AS THIS ANGLE INCREASES, DISABILITY VEILING GLARE EFFECT IS REDUCED
NORMAL LINE OF SIGHT
VERTICAL FT-C AT EYE FROM ALL GLARE SOURCES

Fig. 7 (left). The effect of luminaire mounting height on glare

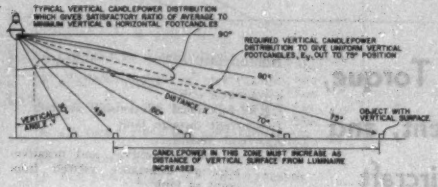

TYPICAL VERTICAL CANDLEPOWER DISTRIBUTION WHICH GIVES SATISFACTORY RATIO OF AVERAGE TO MINIMUM VERTICAL & HORIZONTAL FOOTCANDLES
REQUIRED VERTICAL CANDLEPOWER DISTRIBUTION TO GIVE UNIFORM VERTICAL FOOTCANDLES, E_v, OUT TO 75° POSITION
OBJECT WITH VERTICAL SURFACE
CANDLEPOWER IN THIS ZONE MUST INCREASE AS DISTANCE OF VERTICAL SURFACE FROM LUMINAIRE INCREASES

Fig. 8 (left). Required polar candlepower curve to produce uniform vertical illumination

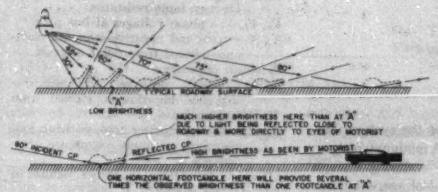

LOW BRIGHTNESS
MUCH HIGHER BRIGHTNESS HERE THAN AT 'A' DUE TO LIGHT BEING REFLECTED GLARE TO ROADWAY & MORE DIRECTLY TO EYES OF MOTORIST
REFLECTED CP
HIGH BRIGHTNESS AS SEEN BY MOTORIST
ONE HORIZONTAL FOOTCANDLE HERE WILL PROVIDE SEVERAL TIMES THE OBSERVED BRIGHTNESS THAN ONE FOOTCANDLE AT 'A'

Fig. 9 (left). The effect of variation in incident angle of light flux on pavement brightness observed by motorist

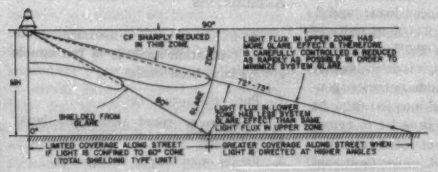

CP SHARPLY REDUCED IN THIS ZONE
LIGHT FLUX IN UPPER ZONE HAS MORE GLARE EFFECT & THEREFORE IS CAREFULLY CONTROLLED & REDUCED AS NEARLY AS POSSIBLE IN ORDER TO MINIMIZE SYSTEM GLARE
SHIELDED FROM GLARE
LIGHT FLUX IN LOWER ZONE HAS LESS SYSTEM GLARE EFFECT THAN SAME LIGHT FLUX IN UPPER ZONE
LIMITED COVERAGE ALONG STREET IF LIGHT IS CONFINED TO 60° CONE (TOTAL SHIELDING TYPE UNIT)
GREATER COVERAGE ALONG STREET WHEN LIGHT IS DIRECTED AT HIGHER ANGLES

Fig. 10 (left). The effect of light flux distribution in the vertical plane on illumination and glare

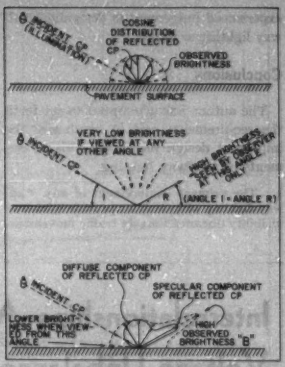

COSINE DISTRIBUTION OF REFLECTED CP
OBSERVED BRIGHTNESS
PAVEMENT SURFACE
VERY LOW BRIGHTNESS IF VIEWED AT ANY OTHER ANGLE
HIGH BRIGHTNESS SEEN BY OBSERVER AT THIS ANGLE ONLY
(ANGLE I = ANGLE R)
DIFFUSE COMPONENT OF REFLECTED CP
SPECULAR COMPONENT OF REFLECTED CP
LOWER BRIGHTNESS WHEN VIEWED FROM THIS ANGLE
HIGH OBSERVED BRIGHTNESS "B"

Fig. 11. Characteristic reflection patterns from different surfaces. Upper view shows diffuse reflection only, middle view shows specular (mirror-like) reflection only, and lower view shows both diffuse and specular reflections. High brightness at "B" is observed when viewing angle is nearly equal to the angle of incident candlepower

practical means their quantitative effect on visibility distance. Therefore, this table does not show any order of merit relationship.

At first glance it becomes obvious that the factors of mounting height and spacing have a major effect on all visual elements. There is no technical problem in meeting the requirements of mounting height and spacing as set forth in reference 4. The effect of mounting height on veiling glare and discomfort glare is shown in Fig. 7.

The candlepower distribution (Figs. 8, 9, and 10) poses the most difficult problem for the design engineer to resolve because of contradictory elements. Those design characteristics which produce high vertical foot-candles for pavement and object brightness also produce both veiling glare and discomfort glare. The solution of this problem becomes a compromise as noted in Fig. 10.

The luminaire area and brightness factor needs no further discussion except with respect to pavement and object brightness. Under wet conditions the street surface becomes more specular and less diffuse. This change in pavement reflection characteristics has a major effect on the roadway brightness patterns (Fig. 11). As the degree of specularity increases, the luminaire size becomes more important. For instance, under extremely wet conditions, smooth flat surfaces become mirrorlike and the observer sees a virtual image of the luminaire. Consequently, the larger the luminaire, the larger the reflected image. Also, the brightness increases as the degree of specularity increases.

The transverse location of the luminaire out over the roadway has a major effect on brightness, veiling glare, and discomfort glare. Increased brightness is produced when the luminaire is placed at or near the center of the roadway, especially as the pavement surface becomes more specular. Generally, this location is confined to residential lighting where silhouette seeing predominates. The effect on glare becomes a function of the angle θ in Fig. 7 except applied in a transverse direction.

All these visual elements can be separately calculated, although glare and pavement brightness calculations are somewhat involved. Measurements can readily be made of vertical and horizontal footcandle intensities. Although more difficult to calculate, brightness measurements can be made of objects and pavement surfaces.

To date, no practical instruments are available for measuring veiling glare and discomfort glare. Until we are able to measure all visual elements and their combined effect on visibility distance and comfort, we must continue to rely on

experienced judgement in designing roadway lighting systems.

Conclusions

The author has attempted to set forth the requirements and to emphasize the important design factors and visual elements of roadway lighting.

1. The primary requirement of any roadway lighting system is to provide adequate visibility distance for safe traffic movement. Comfort, convenience, and appearance are secondary.

2. The visual elements of illumination intensity, pavement and object brightness, uniformity of illumination and brightness, veiling glare, and discomfort glare are effected by luminaire area, candlepower distribution, mounting height, spacing, arrangement and location.

3. There are factors and visual elements which can be calculated, measured, and evaluated as respects visibility distance. There are other factors which can only be evaluated by experienced judgment.

References

1. Tests of Street Illumination. **Preston S. Millar.** *Transactions*, Illuminating Engineering Society, New York, N. Y., vol. XI, June 10, 1916.

2. The Revealing Power of Street Lighting Installations. **J. M. Waldram.** *Ibid.*, London, England, vol. 3, 1938, p. 173.

3. Discomfort Glare at Low Adaptation Levels, Part III—Multiple Sources. **Russell C. Putnam, Keith D. Bower.** *Illuminating Engineering*, New York, N. Y., Apr. 1958.

4. American Standard Practice for Street and Highway Lighting. Illuminating Engineering Society, Feb. 27, 1953, p. 11, Table III.

Interrelationships Among Fault Torque, Voltage Unbalance, Fault Current, and Generator Impedances of Aircraft Electric Power Systems

N. H. JENSEN NONMEMBER AIEE **R. T. SMITH** ASSOCIATE MEMBER AIEE

AMONG the many problems facing those working with aircraft electric power systems is that of satisfying requirements with respect to voltage and current during fault or unbalance conditions, while limiting the torques imposed on the constant-speed drive. It is the purpose of this paper to present a method for obtaining and depicting the interrelationships which exist among these several quantities such that a basis for evaluation in system design may be achieved. The method should help the system engineer and generator designer to arrive at a compromise, giving the best possible system performance in terms of these performance characteristics and generator weight.

The generating system chosen for illustration here consists of four generators rated at 40 kva, 120/208 volts, 0.75-pf (power factor) lag to 0.95-pf lead, 400 cycles per second, 6,000 rpm, and operating either in parallel or isolated. It is assumed that the system frequency is maintained at 400 cycles per second by the use of constant-speed drives.

Paper 58-918, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted March 14, 1958; made available for printing April 17, 1958.

N. H. Jensen is with the General Electric Company, Schenectady, N. Y., and R. T. Smith is with the University of Texas, Austin, Tex.

The specifications of the system illustrated are as follows: A minimum of 4 per-unit current should flow during a single line-to-neutral fault. Voltage unbalance should not exceed 4% during the presence of: 1. an unbalanced load of 33.3% at unity pf with no balanced load; and 2. an unbalanced load of 33.3% at unity pf with a balanced load of 66.7% at unity pf. With respect to torques, it is desirable that the torques developing during faulted conditions and transmitted into the drive be limited to a low value in order to ensure satisfactory operation of the drive. For example, if it is desired, under fault conditions, not to exceed the largest normal steady load torque of the specification, then a value of about 2 per unit, corresponding to twice rated load, would be the upper limit. Recent papers have presented discussions of btoh steady-state[1,2] and transient[1] torques in aircraft electric power systems.

Calculations of fault current, voltage unbalance, and drive torques were made using methods already presented in the literature; see references. All these quantities were obtained as functions of the alternator impedances and plotted on common axes. By normalizing the various quantities with respect to the specified limits, the effect of changes in machine impedances on these factors may be observed very clearly. Computations were carried out on a digital computer.

Nomenclature

a = operator $e^{j120°}$
f, g, h = functional relationship symbols
I_F = current through fault
R_F = fault resistance
R_0, R_1, R_2 = zero-, positive-, and negative-sequence resistance viewed from point of fault
T_{max} = maximum steady-state drive torque for any fault resistance
V_a, V_b, V_c = phase voltages at bus point
X_0, X_2 = zero- and negative-sequence reactance viewed from point of fault
Z_F = fault impedance
Z_0, Z_1, Z_2 = zero-, positive-, and negative-sequence impedances viewed from fault point
Z_{0G}, Z_{1G}, Z_{2G} = generator zero, positive, and negative per-unit sequence impedance
ΔU = per-cent voltage unbalance, defined in *MIL-G-6099A*[3]

Assumptions

The following general assumptions for the co-ordination analysis of the sample system have been made:

1. Pure resistance in fault.
2. Synchronous speed maintained during all conditions.
3. Faults occur on the system bus.
4. Maximum generator temperature of 210 C (degrees centigrade).
5. Maximum feeder temperature of 110 C.
6. Minimum machine and feeder temperatures of 25 C during any load or fault condition.
7. Class *C* excitation, with static exciter and highest phase sensing.
8. Feeder impedances:
   $Z_1 = Z_2 = 0.026 + j0.020$ per unit.
   $Z_0 = 0.044 + j0.123$ per unit.

For the most part these assumptions represent rather extreme conditions so far as the factors under consideration are concerned.

Fault Current

In the load feeders of the distribution system, short-circuit protection is

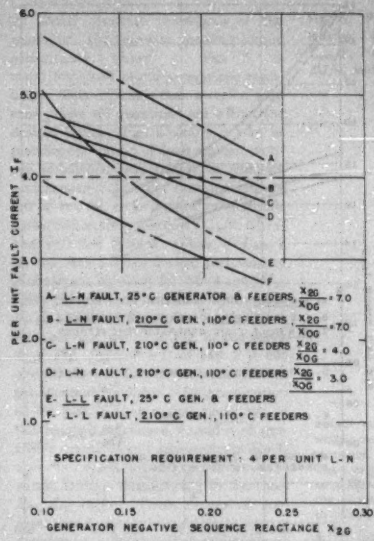

PER UNIT FAULT CURRENT I_F
A- L-N FAULT, 25°C GENERATOR & FEEDERS, $\frac{X_{2G}}{X_{OG}}$ = 7.0
B- L-N FAULT, 210°C GEN., 110°C FEEDERS $\frac{X_{2G}}{X_{OG}}$ = 7.0
C- L-N FAULT, 210°C GEN., 110°C C FEEDERS $\frac{X_{2G}}{X_{OG}}$ = 4.0
D- L-N FAULT, 210°C GEN., 110°C C FEEDERS $\frac{X_{2G}}{X_{OG}}$ = 3.0
E- L-L FAULT, 25° C GEN. & FEEDERS
F- L-L FAULT, 210° C GEN., 110° C FEEDERS
SPECIFICATION REQUIREMENT: 4 PER UNIT L-N
GENERATOR NEGATIVE SEQUENCE REACTANCE X_{2G}

Fig. 1 (left). Fault current for line-to-neutral and line-to-line faults, no load on system during fault

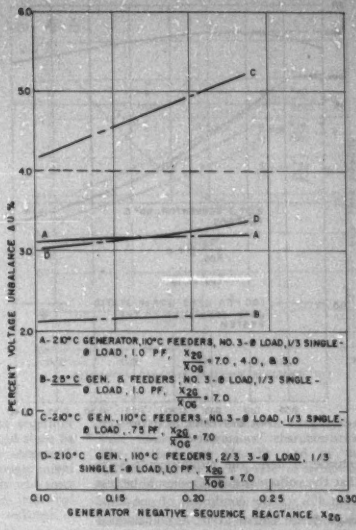

PERCENT VOLTAGE UNBALANCE AU %
A- 210°C GENERATOR, 110°C FEEDERS, NO. 3-Ø LOAD, 1/3 SINGLE-Ø LOAD, 1.0 PF, $\frac{X_{2G}}{X_{OG}}$ = 7.0, 4.0, & 3.0
B- 25°C GEN. & FEEDERS, NO. 3-Ø LOAD, 1/3 SINGLE-Ø LOAD, 1.0 PF, $\frac{X_{2G}}{X_{OG}}$ = 7.0
C- 210°C GEN., 110°C FEEDERS, NO. 3-Ø LOAD, 1/3 SINGLE-Ø LOAD, .75 PF, $\frac{X_{2G}}{X_{OG}}$ = 7.0
D- 210°C GEN., 110°C FEEDERS, 2/3 3-Ø LOAD, 1/3 SINGLE-Ø LOAD, 1.0 PF, $\frac{X_{2G}}{X_{OG}}$ = 7.0
GENERATOR NEGATIVE SEQUENCE REACTANCE X_{2G}

Fig. 2 (right). Voltage unbalance

achieved through the use of overcurrent protective devices. Short-circuit current will cause the faulted line protective elements to act, thereby isolating the fault. This protection is effective only if the system has sufficient short-circuit current capacity. The minimum requirements for this capacity are usually set forth in the system specification. The requirements for this system are 3 per-unit 3-phase fault current for 5 seconds and 4 per-unit line-to-neutral fault current for 5 seconds, both for no fault impedance. These values should obtain at all operating conditions in order to assure reliable performance, which means that the most severe conditions need be considered.

Three-phase fault current capacity is limited primarily by the exciter. Thus, this condition is satisfied by including it as an exciter design criterion. During unbalanced faults the exciter will, during most operating conditions, have sufficient capacity to maintain one of the phase voltages at rated voltage, in the case of highest phase sensing. Thus, the unbalanced fault current will usually be limited by regulator action rather than exciter capability. Then if the sequence constants of the system are known, it is relatively easy to formulate a relationship between the fault currents and phase voltages. By making use of the fact that the highest phase voltage is maintained (at 1.0 per unit), the unbalanced fault current may be written as a function of the system zero- and negative-sequence impedances,[4,5] functionally,

$$I_F = f(R_2, X_2, R_0, X_0) \qquad (1)$$

In order to present conservative results, calculations have been made for the case of no load during fault. In the sample system lower currents will flow in the generator lines under this condition than if a load is connected to the system during the fault.

Fig. 1 shows unbalanced fault current as a function of generator negative-sequence reactance. Curves *A* and *B* show the significant effects of temperature upon line-to-neutral fault current, while curves *E* and *F* show the same influence for line-to-line faults. Curves *B*, *C*, and *D* show the influence of variations in generator zero-sequence reactance. Changes in generator negative- and zero-sequence resistance could also be considered. However, experience has shown that, due to design considerations, variations in these quantities at a given temperature are small, and further that total system quantities will be affected to an even smaller degree. It is thought, therefore, that the values used here are representative.

Voltage Unbalance

The maximum permissible unbalance is stated in the specification for the electrical system. For example, in the system

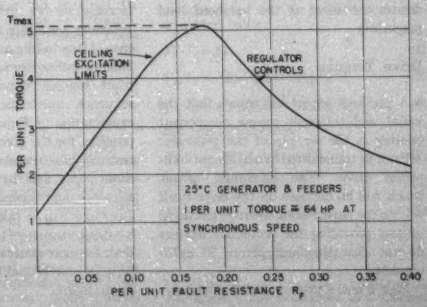

T_{max}
CEILING EXCITATION LIMITS
REGULATOR CONTROLS
PER UNIT TORQUE
25°C GENERATOR & FEEDERS
1 PER UNIT TORQUE = 64 HP AT SYNCHRONOUS SPEED
PER UNIT FAULT RESISTANCE R_F

Fig. 3. Steady-state torque of 40-kva generating system with balanced 3-phase fault resistance R_F

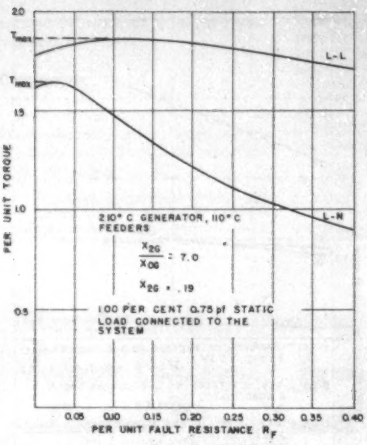

PER UNIT TORQUE
PER UNIT FAULT RESISTANCE R_F
L-L
L-N
210° C GENERATOR, 110° C FEEDERS
$\frac{X_{2G}}{X_{0G}} = 7.0$
$X_{2G} = .19$
100 PER CENT 0.75 pf STATIC LOAD CONNECTED TO THE SYSTEM

Fig. 4 (left). Steady-state fault torque (line-to-line, line-to-neutral) of 40-kva generating system

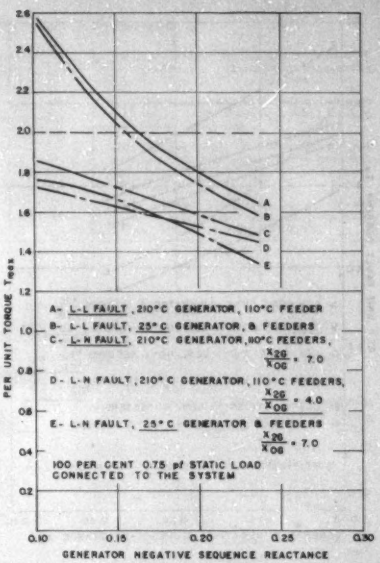

PER UNIT TORQUE T_{MAX}
GENERATOR NEGATIVE SEQUENCE REACTANCE
A- L-L FAULT, 210° C GENERATOR, 110° C FEEDER
B- L-L FAULT, 25° C GENERATOR, & FEEDERS
C- L-N FAULT, 210° C GENERATOR, 80° C FEEDERS, $\frac{X_{2G}}{X_{0G}} = 7.0$
D- L-N FAULT, 210° C GENERATOR, 110° C FEEDERS, $\frac{X_{2G}}{X_{0G}} = 4.0$
E- L-N FAULT, 25° C GENERATOR & FEEDERS $\frac{X_{2G}}{X_{0G}} = 7.0$
100 PER CENT 0.75 pf STATIC LOAD CONNECTED TO THE SYSTEM

Fig. 5 (right). Maximum fault torque for line-to-neutral and line - to - line faults. Rated load: 0.75 pf connected to system during the fault

considered here, the requirements were that the voltage unbalance should be less than 4% for the conditions of one-third single-phase load at 1.0 pf, at no balanced load, at one-third balanced load, and at two-thirds balanced load.

For given load conditions it is possible to express the per-cent voltage unbalance, ΔU, as a function of the generating system sequence constants.[5,6,7] Thus, as was the case for fault currents,

$$\Delta U = g(R_2, X_2, R_0, X_0) \qquad (2)$$

Fig. 2 shows a plot of voltage unbalance versus generator negative-sequence reactance for different load and temperature conditions. Curve *A* is for 1.0-pf single-phase load with zero balanced load. For this condition variations in X_{0G} were found to have a negligible effect so that a single plot is shown for X_{2G}/X_{0G} ratios of 3.0, 4.0, and 7.0. Curve *B* shows the effect of temperature. Curve *C* depicts the influence of pf of the unbalanced load on voltage unbalance, while curve *D* indicates the effect of the balanced load magnitude.

Drive Torques

A previous paper[1] has shown that the initial high-transient torque component existing in the air gap of the generator will not be transmitted to a hydraulic ball-piston drive. The maximum torque which will be imposed on the drive itself will be the steady-state average electrical torque, providing the protective devices do not clear the fault before 50 milliseconds.

Figs. 3 and 4 present curves of steady-state drive torques during 3-phase and unbalanced faults, respectively. At the present time very little is known concerning the magnitude of fault impedances on a-c systems. Therefore, to be conservative, the largest possible torque T_{max} for any fault resistance should be considered.

It would be desirable, of course, if the drive were able to carry any torques which might develop, without speed reduction or damage. Typical specification load requirements are 150% load for 5 minutes and 200% load for 5 seconds. Examination of Fig. 3 shows that the torques resulting from 3-phase faults exceed 200% by a considerable amount. To design the drive to carry the maximum 3-phase fault torques would result in a large weight increase in the drive and possibly in accessory gear train and pad also. In the case of line-to-line and line-to-neutral unbalance faults the torques may be somewhat smaller, depending upon system sequence constants. It is seen (from examination of Figs. 4 and 5) that it is possible, for the given load conditions, to restrict these unbalanced torques to less than 2 per unit by proper selection of sequence impedances. Therefore, $T_{max} = 2$ per unit will be taken as the upper limit for drive torque. It must be emphasized that, for more general load conditions, the torques can probably not be restricted to this figure.

The magnitude of the torques is a function of the fault currents and the resistance present in the circuit. Since a relationship has already been established for determining the unbalanced fault currents, an expression for unbalanced fault drive torques can also be written. By plotting these torques as a function of fault resistance (see Fig. 4), the maximum drive torque may be found directly. Therefore, a functional relationship exists such that

$$T_{max} = h(R_1, R_2, X_2, R_t, X_t) \qquad (3)$$

Fig. 5 gives a plot of maximum steady-state average torque imposed on the drive as a function of the generator negative-sequence reactance for various temperatures, unbalanced faults, and X_{2G}/X_{0G} ratios, for highest off-phase voltage maintained at 1.0 per unit. Under certain of these conditions, it is possible that a reduction in torque would occur as a result of excitation limitations, which have not been taken into account in this sample analysis. However, in this way these results are made somewhat conservative, which can be verified by excitation calculations.

Interrelationships

Figs. 1 and 2 show that the fault current and voltage unbalance performance is best

for small sequence reactances of the generator. Contrariwise, the fault torque performance, shown in Fig. 5, is worst for small negative-sequence reactance. This indicates that a region may exist where all quantities are within the desired limits. Fig. 6 shows a way of plotting the curves of fault current, voltage unbalance, and torque to compare performance. In order that ordinates of 1.0 or less be an indication of desirable performance, the quantities $T_{max}/2$, $\Delta U/4$, and $4/I_F$ are plotted. Thus, if $T_{max}/2$ be less than 1.0, the desired performance is obtained; similarly for $\Delta U/4$ and $4/I_F$. Data plotted were obtained from curves *B*, *C*, and *D* of Fig. 1, the combination solid curve *A* and *D* of Fig. 2, and curve *A* of Fig. 5. Curve *C* of Fig. 2 is not plotted since it does not correspond to a specification condition. By examination of the plot it is seen that there are, in this example, regions *a*, *b*, and *c*, of generator negative- and zero-sequence reactance in which system performance is within limits, corresponding to the three X_{20}/X_{00} ratios. These ranges of generator constants can be used by the system designer as a guide during design stages or as a convenient check on how a given generator will affect system performance with respect to fault currents, voltage unbalance, and torques. In this regard it must be noted that for certain postulated system requirements of fault current, voltage unbalance, and drive torque, the acceptable regions of generator sequence reactance could become impractically small for the generator designer to meet. Thus, it is clear that torque, voltage unbalance, and current requirements must be compatible with realistic generator design, for otherwise an excessively heavy system may result. This illustrates the importance of the close co-operation between the generator designer and the systems engineer.

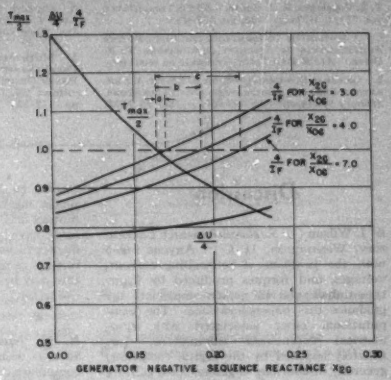


Fig. 6. Performance characteristics of fault torque $T_{max}/2$, voltage unbalance $\Delta U/4$, and fault current $4/I_F$

Discussion

1. The general assumptions were chosen either to be representative of the system operation or to produce conservative results. The method is adaptable, however, to a wide range of systems working under various operating conditions.

2. The selection of extreme temperature conditions was based upon a study of previous systems of similar design and environment.

3. If the drive design criterion is the 2 per-unit maximum load, then the main weight variables of the system are the generator and the feeders. A single feeder cable between generator and bus was considered in the sample analysis given here. Analysis for two, three, or four parallel cables could also be made. Although cable weight would likely increase, the generator weight could probably be reduced since its sequence constants could be increased. Thus, an over-all lower system weight might be achieved. This is particularly true since the cooling of multiple cables is improved so that copper weight may be reduced.

4. A previous paper[1] indicates that a hydraulic drive of the ball-piston type does not sense the high-transient torques developed in the generator air gap which persist for less than approximately 25 milliseconds. If a fault occurs in a section of the electrical system which is protected by a differential relay, the fault will normally be cleared in less than 25 milliseconds. Therefore, even the full steady-state torque will not be imposed on the drive. Similarly, if the fault occurs behind an overcurrent device which will clear in less than 25 milliseconds, the torque imposed on the drive will be less than the steady-state torque produced by the generator.

5. In the example given, fault torques were calculated on the basis of rated static load impedance on the system during fault. Loads such as motors and transformer rectifiers may have low zero- and/or negative-sequence impedances. These loads can cause the unbalanced fault torques to increase beyond the values shown in the sample analysis. Resulting torques might even be so high that the penalties for obtaining co-ordination may be excessive, unless special precautions are taken. These could be excitation control to limit the torques to tolerable values or disconnection of major motor loads during a fault condition.

Conclusions

1. A method is presented for depicting the interrelationships which exist among fault current, voltage unbalance, and fault torque imposed on the constant-speed drive, as a function of the generator sequence impedances. This procedure may be used to determine system constants so that fault current and voltage unbalance meet specification requirements. At the same time, unbalanced fault torques imposed on the drive may be co-ordinated with these performance characteristics. The method may also be used by the system designer as a guide to achieving minimum over-all system weight.

2. It is seen that efforts to improve voltage unbalance and increase fault currents will, at the same time, result in increased drive torques. Therefore, the method described can help in realistically appraising the necessary compromises involved in this area.

3. While the method has been illustrated for a specific system, with balanced static loads, it is general enough to allow many types of loads such as motor and transformer rectifiers to be handled. It should be emphasized that load conditions may have considerable effect upon over-all results.

Appendix

Equations for voltage unbalance for 3-phase systems are given in a comprehensive paper by Wilson and Gardner.[6] Hardman[7] presents voltage unbalance as a function of generator impedances. Fault torque expressions are given in a paper by Caldwell, Coppinger, and Smith.[3]

For fault current calculations the following equations may be used.[4,5] For line-to-neutral faults:

$$V_a/I_F = Z_F \quad (4)$$

$$V_b/I_F = a^2Z_F - j(Z_1 - aZ_2)/\sqrt{3} \quad (5)$$

$$V_c/I_F = aZ_F + j(Z_1 - a^2Z_2)/\sqrt{3} \quad (6)$$

For line-to-line faults:

$$V_a/I_F = -j(2Z_1 + Z_F)/\sqrt{3} \quad (7)$$

$$V_b/I_F = j(Z_1 - a^2Z_F/\sqrt{3} \quad (8)$$

$$V_c/I_F = j(Z_1 - aZ_F)/\sqrt{3} \quad (9)$$

References

1. Co-ordination of Hydraulic Transmissions and Aircraft A-C Generators, S. C. Caldwell,

T. E. Coppinger, R. T. Smith. *AIEE Transactions*, vol. 77, pt. II, Nov. 1958, pp. 377–88.

2. A Method of Determining the Input Power Requirements of Aircraft Generators, T. B. Owen. *AIEE DP-57-488* (available on request).

3. General Specification for Aircraft Alternating Current Generators and Regulators. *Military Specification MIL-G-6099A*, Mar. 25, 1957.

4. Circuit Analysis of A-C Power Systems, Vol. I (book), E. Clarke. John Wiley & Sons, Inc., New York, N. Y., 1943.

5. Methods for Prediction of Steady-State Performance for Unbalanced Regulated 3-Phase Generators, B. J. Wilson. *AIEE Transactions*, vol. 72, pt. II, 1953 (Jan. 1954 section), pp. 413–22.

6. Calculations on Voltage Unbalance for 3-Phase Synchronous Systems, B. J. Wilson, W. K. Gardner. *Ibid.*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 426–37.

7. The Effect of Machine Impedances on the Voltage Unbalance of 3-Phase Synchronous Generators, T. F. Hardman. *Ibid.*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 467–71.

Discussion

B. J. Wilson (U. S. Naval Research Laboratory, Washington, D. C.): Anyone faced with the problem of reconciling current, voltages, and torques produced by fault-type unbalances will greatly appreciate the guidance this paper furnishes. The computational labor associated with symmetrical component calculations, fortunately lessened by the digital computer, generally discourages engineers from gaining what we might term the topological picture presented in the paper. Although the reactance values are limited with respect to the range considered, they cover the range encountered in modern aircraft practice.

Although the factors of voltage, current, and torque are compromised for "best" fault conditions, it must be pointed out that they are not reconciled with two other important factors—operation under less severe unbalance, and effect on the operation of load equipment such as induction motors. For example, the increase in negative-sequence reactance, to lessen fault torque magnitude, increases voltage unbalance at a given load unbalance. This results in degraded performance of polyphase machinery in general, but notably motors and the generator itself. It would also have been instructive to indicate discrepancies between analytical and experimental results, although in a broad sense this may be considered outside the scope of the paper.

N. H. Jensen and **R. T. Smith**: The authors wish to thank Mr. Wilson for his comments. We agree that it is important to recognize that voltage unbalance does affect motor performance. It was assumed in our presentation that so long as unbalance was within specifications, motor performance need not be considered.

Test and calculated values of fault current and voltage unbalance have generally agreed well with one another; i.e., discrepancies of less than 5% can be expected; see Table I. However, test data on fault torques have shown that calculated values consistently were about 15% too low. We think this discrepancy is caused principally by the harmonic losses together with friction, windage, eddy-current, and other losses not accounted for in the calculations. Therefore, in order to bring the calculations in closer agreement with actual values, a correction factor of 1.15 has been used in calculating the torques shown in Figs. 3 through 6. Subsequent comparisons of test and calculated torques (using this factor) have shown excellent agreement.

Table I. Test Over Calculated Results

Item	Test/Calculated Ratio			
Systems rating, kva...	20...	22...	40...	60
Negative-sequence reactance.....	0.99...	1.07...	0.91...	1.00
Fault current, line-to-neutral.....	0.98...	0.96...	1.03...	1.02
Fault current, line-to-line.....	1.03...	0.95...	1.08...	1.00
Voltage unbalance....	1.02...	1.04...	1.10...	1.05

The use of digital computers in the design of aircraft generators has greatly improved the accuracy of predicted impedances. Consequently, good agreement exists between analytical and experimental results.[1]

Reference

1. Computer Analysis of A-C Aircraft Generators, J. R. M. Alger, E. F. Magnusson, J. T. Duane, R. T. Smith. *AIEE Transactions*, vol. 77, pt. II, Nov. 1958, pp. 394-99.

Mathematical Model and Procedure for Arc-Back Current Calculations for Power Rectifiers

H. P. FULLERTON
NONMEMBER AIEE

J. TENO
ASSOCIATE MEMBER AIEE

THE anode and cathode circuit breaker fault currents experienced by mercury-arc rectifier equipment when an arc back occurs have generally been estimated by superposition of the results of greatly simplified calculations of individual a-c and d-c transient states. Such calculations have the merit of being easy and rapid to make. In some instances however the large differences between such estimates and actual tests have indicated a need for a more precise method of predicting the arc-back currents in rectifier circuits. This need has been particularly pressing in the case of circuits having large connected capacity, for as systems get larger the accurate prediction of the magnitude of arc-back currents to be handled becomes increasingly important in the design and selection of components.

This paper discusses a more complete mathematical model proposed for arc-back calculations, the evaluation of the constants required in carrying out such calculations, and a brief description of the methods used to carry out these calculations on a medium-size electronic digital computer.

The general comparison of test results with those computed by the methods described herein are given in another paper.[1] A sample set of results is included in this paper to illustrate the degree of agreement obtained.

The model was selected not only to permit a general analysis of the transient behavior of the mercury-arc rectifier but also to investigate a number of specific questions, some of which are:

1. What is the contribution of the normal (unfaulted) Y to the arc-back current?
2. How does interphase saturation affect the arc-back current?
3. Is it possible to include the breaker's current-limiting action in the analysis and,

Paper 58-226, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2–7, 1958. Manuscript submitted November 4, 1957; made available for printing June 5, 1958.

H. P. Fullerton is with the General Electric Company, Philadelphia, Pa., and J. Teno is with Lehigh University, Bethlehem, Pa.

The authors gratefully acknowledge the assistance of A. Schmidt, Jr., and C. H. Titus in the formulation of the model and evaluation of the system constants, and G. K. Carter in the adaptation of the problem for computer solution.

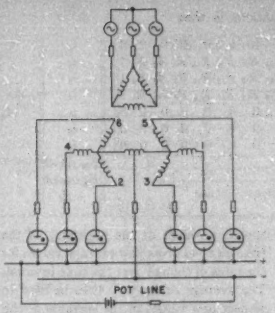


Fig. 1. Single double-Y rectifier system

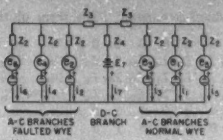


Fig. 2. Rectifier equivalent circuit

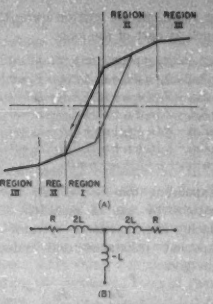


Fig. 3. Representation of interphase transformer

A—Linearized magnetization curve
B—Equivalent circuit, assumed K=1.0 or L=M

if so, what is the effect of the breaker action on the arc-back current under various conditions?

The complexity of the equations required to answer such questions have made it virtually impossible to analyze arc-back currents using elementary computational methods. The advent of the electronic computer has made the analysis of this problem practicable. In the work done by the authors, an International Business Machines Corporation *650* computer was used to obtain the solution to the set of circuit equations that were derived. The analysis is in itself rigorous so that not only can the questions previously outlined be easily answered but exact answers to other questions concerning the system can be easily and quickly obtained. Results of such an analysis are extremely accurate as is demonstrated by a comparison of test and calculated data. Field tests which led to the development of this procedure for calculating arc-back current were conducted at the Rockdale (Texas) Plant of the Aluminum Company of America. The model developed herein reflects the characteristics of that system but is believed to be readily applicable to many rectifier systems. Although developed for determination of breaker duty, the information available from the computation is equally applicable to the design or selection of suitable ratings of other system components.

Nomenclature

$D = d/dt$, differential operator
e = instantaneous voltage
E_{ac} = voltage on secondary a-c bus, line to line, rms
$(E_{ac})_M$ = secondary bus, maximum value of voltage, line to line
E_{dc} = d-c bus voltage, open circuit
E_M = equivalent generator voltage, maximum, see equation 7
$E_7 = K_3(E_{ac})_M$, equivalent d-c branch source voltage
$i_1, i_2, \ldots, i_6$ = rectifier a-c branch currents
i_7 = d-c branch current
$i_{ipt} = (i_2+i_4+i_6)-(i_1+i_3+i_5)$ = net magnetizing current, interphase transformer
ipt = interphase transformer (used as subscript)
I_{PL} = potline current, total
K = coefficient of coupling, interphase transformer
K_1 = system constant for normally operating rectifiers which relates magnitude of assumed sinusoidal alternating current supplied to d-c
K_3 = system constant for normally operating rectifiers which relates d-c output voltage to a-c bus voltage supplied
L = inductance
M = mutual inductance
R = resistance
R_{COM} = equivalent resistance to account for commutation drop in normal rectifier units
t = time
X = inductive reactance
$Z = R+L(d/dt)$, impedance operator
Z_A = impedance of supply system to secondary bus which is common feed for all rectifier transformers connected to single potline
Z_L = impedance of high-voltage system excluding leakage impedance of transformers
Z_T = leakage impedance of rectifier transformer
Z_1 = impedance of high-voltage system from neutral of equivalent source to and including transformer leakage impedance
Z_2 = impedance of low-voltage system from transformer to unit cathode bus
Z_3 = impedance associated with interphase transformer
Z_4 = combined impedance of equivalent d-c system and mutual reactance of interphase transformer
α = phase-shift angle between rectifier transformers
$\theta = \omega t$
ϕ = initial phase angle for generator voltage
ω = angular frequency of applied voltage

Single-Rectifier System

A simple system in which a potline is supplied by a single double-Y rectifier is shown in Fig. 1. A system in which a common potline bus is supplied by rectifiers operating in parallel is considered later. The concept of a single-rectifier system serves to introduce the methods used for computer solution. The system of Fig. 1 may be reduced to the simple equivalent circuit shown in Fig. 2 in which two or three a-c branches and the d-c brauch are connected at any time, the remaining a-c branches being open. The generalized impedance operator $[R+L(d/dt)]$ obtained by a series combination of the phase resistance and phase inductance is designated by the symbol Z and is assumed to be identical for all phases. It is further assumed that the high-voltage side of the system is excited by a balanced set of sinusoidal voltages. The a-c branch impedances Z_2 are those for the low-voltage side of the system of Fig. 1, and the branch voltages e_1, e_2, ..., e_6 are the instantaneous values of the available voltage across the windings of the low side of the rectifier transformer. Expressions for these branch voltages are given in Appendix I.

Rectifier tubes are represented by a constant resistance included in each branch. Additional linearization of the circuit is obtained by representing the variable inductance of the saturating interphase transformer by a magnetization curve of the form shown in Fig. 3(A). The equivalent circuit for the interphase transformer is shown in Fig. 3(B). The evaluation of the impedances for these various parts of the circuit is discussed in detail later.

A master set of equations may be written for the circuit of Fig. 2. From this set the appropriate equations may be selected based on the branches conducting for a particular condition of operation of the circuit. The following

loop equations are written using the d-c branch as part of the loop.

$$
\begin{aligned}
&e_1 - Z_3(i_1+i_3+i_4) + Z_4 i_7 - E_7 = 0\\
&e_2 - Z_3(i_2+i_4+i_5) + Z_4 i_7 - E_7 = 0\\
&e_3 - Z_3(i_3+i_5+i_6) + Z_4 i_7 - E_7 = 0\\
&e_4 - Z_3(i_4+i_5+i_6) + Z_4 i_7 - E_7 = 0\\
&e_5 - Z_3(i_5+i_6+i_4) + Z_4 i_7 - E_7 = 0\\
&e_6 - Z_3(i_6+i_4+i_5) + Z_4 i_7 - E_7 = 0\\
&i_1+i_2+i_3+i_4+i_5+i_6+i_7 = 0
\end{aligned}
\qquad (1)
$$

Expanding the first of the foregoing equations by use of equation 32 and replacing the impedance operator by its component resistance and inductance terms give

$$
\begin{aligned}
&(L_1+L_3+L_3-L_L)\frac{di_1}{dt} + L_L\frac{di_2}{dt} +\\
&\quad (L_3-L_L)\frac{di_3}{dt} - (L_1-L_L)\frac{di_4}{dt} +\\
&\quad (L_3-L_L)\frac{di_5}{dt} + L_L\frac{di_6}{dt} - L_4\frac{di_7}{dt}\\
&\quad = -(R_1+R_3+R_3-R_L)i_1 - R_L i_2 -\\
&\quad (R_3-R_L)i_3 + (R_1-R_L)i_4 - (R_3-R_L)i_5 -\\
&\quad R_L i_6 + R_4 i_7 + E_M \cos\phi_1 \sin\omega t +\\
&\qquad E_M \sin\phi_1 \cos\omega t - E_7
\end{aligned}
\qquad (2)
$$

The other loop equations have a similar form. Equation 1 may then be arranged in matrix form, separating the constant and variable terms. In doing this it is found desirable to substitute its derivative for the last equation in equation 1.

$$
\begin{Vmatrix}
L_{11} & L_{12} & L_{13} & L_{14} & L_{15} & L_{12} & L_6\\
L_{12} & L_{11} & L_{12} & L_{14} & L_{14} & L_{12} & L_6\\
L_{13} & L_{12} & L_{11} & L_{12} & L_{13} & L_{14} & L_6\\
L_{14} & L_{12} & L_{12} & L_{11} & L_{12} & L_{13} & L_6\\
L_{13} & L_{14} & L_{13} & L_{12} & L_{11} & L_{12} & L_6\\
L_{13} & L_{13} & L_{14} & L_{13} & L_{12} & L_{11} & L_6\\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{Vmatrix}
\cdot
\begin{Vmatrix} Di_1\\ Di_2\\ Di_3\\ Di_4\\ Di_5\\ Di_6\\ Di_7 \end{Vmatrix}
=
\begin{Vmatrix}
R_{11} & R_{12} & R_{13} & R_{14} & R_{13} & R_{12} & R_6 & A_1 & B_1 & 1\\
R_{12} & R_{11} & R_{12} & R_{13} & R_{14} & R_{13} & R_6 & A_2 & B_2 & 1\\
R_{13} & R_{12} & R_{11} & R_{12} & R_{13} & R_{14} & R_6 & A_3 & B_3 & 1\\
R_{14} & R_{13} & R_{12} & R_{11} & R_{12} & R_{13} & R_6 & A_4 & B_4 & 1\\
R_{13} & R_{14} & R_{13} & R_{12} & R_{11} & R_{12} & R_6 & A_5 & B_5 & 1\\
R_{12} & R_{13} & R_{14} & R_{13} & R_{12} & R_{11} & R_6 & A_6 & B_6 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{Vmatrix}
\cdot
\begin{Vmatrix} i_1\\ i_2\\ i_3\\ i_4\\ i_5\\ i_6\\ i_7\\ E_M \sin\omega t\\ E_M \cos\omega t\\ -E_7 \end{Vmatrix}
\qquad (3)
$$

where

$$
\begin{aligned}
&L_{11} = L_1 + L_3 + L_3 - L_L\\
&L_{12} = L_L\\
&L_{13} = L_3 - L_L\\
&L_{14} = -(L_1 - L_L)\\
&L_6 = -L_4\\
&R_{11} = -(R_1 + R_3 + R_3 - R_L)\\
&R_{12} = -R_L\\
&R_{13} = -(R_3 - R_L)\\
&R_{14} = R_1 - R_L\\
&R_6 = R_4\\
&A_n = \cos\phi_n\\
&B_n = \sin\phi_n\\
&Di_n = \frac{di_n}{dt}
\end{aligned}
$$

While these are the general equations, at any particular time many of the branches are nonconducting because of the open-circuited condition of their tubes, and such branches are represented in the matrix form by an equation $Di_n = 0$. As an example, consider the case when branches 2, 3, and 4 are conducting current. The form of the equations is then

$$
\begin{Vmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0\\
L_{12} & L_{11} & L_{13} & L_{13} & L_{14} & L_{13} & L_6\\
L_{13} & L_{12} & L_{11} & L_{12} & L_{13} & L_{14} & L_6\\
L_{14} & L_{13} & L_{13} & L_{11} & L_{12} & L_{13} & L_6\\
0 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0\\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{Vmatrix}
\cdot
\begin{Vmatrix} Di_1\\ Di_2\\ Di_3\\ Di_4\\ Di_5\\ Di_6\\ Di_7 \end{Vmatrix}
=
\begin{Vmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
R_{12} & R_{11} & R_{13} & R_{13} & R_{14} & R_{13} & R_6 & A_2 & B_2 & 1\\
R_{13} & R_{13} & R_{11} & R_{13} & R_{13} & R_{14} & R_6 & A_3 & B_3 & 1\\
R_{14} & R_{13} & R_{13} & R_{11} & R_{13} & R_{13} & R_6 & A_4 & B_4 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{Vmatrix}
\cdot
\begin{Vmatrix} i_1\\ i_2\\ i_3\\ i_4\\ i_5\\ i_6\\ i_7\\ E_M \sin\omega t\\ E_M \cos\omega t\\ -E_7 \end{Vmatrix}
\qquad (4)
$$

In order to agree with the notation of the Rockdale tests, the arc back is assumed to occur in branch 2 of the rectifier. The problem solution starts at the beginning of conduction in branch 4 as indicated in Fig. 4. Equation 4 applies at this time.

The equations are now in a systematic form which lends itself to computer solution. A general program may be written, controllable for different branches conducting, which will insert the proper values of the coefficient matrices of equation 4 at each stage of the computer solution. The next step is to solve these seven simultaneous equations for seven new equations of the form

$$
\frac{di_n}{dt} = f_n(t, i_1, i_2 \ldots i_7) \qquad (5)
$$

Then, at a particular time, having the values of the currents in the respective branches, the corresponding rates of change may be calculated.

With the derivatives evaluated it is possible to proceed with a step-by-step integration process often called "the modified Euler method without iteration."[2] Starting at time zero with known (or assumed) initial values of current in branches 2, 3, and 7, rates of change of the respective currents are readily computed. Having established these, projections through a time interval Δt may be made to yield a new value for each current, which will be called the "predicted" value. These in turn may be substituted in the f_n functions to obtain estimates of the rates of change at the end of the time interval. Averaging each of these last with the corresponding values at the beginning of the time interval gives new average values for the rates of change over the time interval. The average values may then be used to project new values of the currents at the end of the time interval, which will be called the "corrected" values. A measure of the success of the process may be obtained by comparing the predicted and corrected values for each branch. In general, the integration is progressing smoothly and accurately when this difference is small. A relatively large difference indicates that an excessively large Δt is being used and a reduction in Δt is required to bring the "error" within the permissible range. The procedure used by the authors automatically makes a recalculation over a reduced time interval whenever the error exceeds a specified amount.

Successive cycles of computation advance the calculations on time until a point is reached at which there is a change in circuit conditions. This may be a switching in of a new branch, a switching out of a branch in which commutation has been completed, or a change in the inductance level of the interphase transformer due to a change in the degree of saturation. By testing the appropriate circuit values at the end of each time interval, it may be determined when any such change in circuit condition is indicated. When such a point has been passed, the program sets into motion an interpolation routine which, by linear interpolation, determines the time at which the test point was passed and adjusts all current values

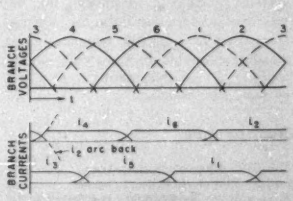


Fig. 4. Rectifier steady-state current and voltage relations

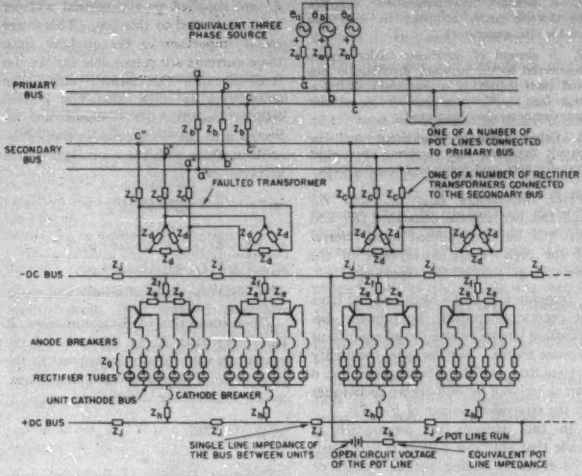

EQUIVALENT THREE PHASE SOURCE
PRIMARY BUS
SECONDARY BUS
FAULTED TRANSFORMER
ONE OF A NUMBER OF POT LINES CONNECTED TO PRIMARY BUS
ONE OF A NUMBER OF RECTIFIER TRANSFORMERS CONNECTED TO THE SECONDARY BUS
-DC BUS
ANODE BREAKERS
RECTIFIER TUBES
UNIT CATHODE BUS
CATHODE BREAKER
+DC BUS
POT LINE RUN
SINGLE LINE IMPEDANCE OF THE BUS BETWEEN UNITS
OPEN CIRCUIT VOLTAGE OF THE POT LINE
EQUIVALENT POT LINE IMPEDANCE

Fig. 5. Typical potline rectifier system

accordingly. The routine then assembles the coefficients for the appropriate circuit equations for the next circuit condition, solves for the coefficients of the appropriate f_n equations as before, and starts calculating again in the same fashion but with the new equations for the rates of change of currents.

This process may be repeated through each new circuit condition until the first peak of the fault current has been passed or possibilities provided in the program have been exhausted.

Multiple Rectifiers Feeding a Common Bus

Having developed the general lines of the solution in terms of a simple single-rectifier system, the effect of paralleling rectifiers supplying the same d-c bus is now considered. A typical system of this type is shown in Fig. 5. The sources have been reduced to a single equivalent 3-phase source. The effect of the regulation of any other load on the primary bus is neglected. The equivalent values of impedance and source voltage to be used in the d-c branch of Fig. 2 are more difficult to evaluate for this case. It now represents a lumping of all loads and d-c sources connected to the d-c bus. (A matrix method for accomplishing this is derived in Appendix III.) The increased line currents resulting from the arc back will cause a drop in the secondary bus voltage, and the d-c output voltage of the other rectifiers will dip. To reflect this, the value of E_T is held proportional to the voltage at the secondary bus and no longer is constant as in the case of a single rectifier.

The branch voltages $e_1, e_2, \ldots, e_6$ are reduced by the amount of the voltage drop contributed by the a-c line current drawn by the paralleling rectifiers through the impedance between the source and the secondary bus. This additional voltage drop is

$$\Delta e_n = 3Z_A K_1 (I_{PL} + i_T) \sin(\omega t + \phi_n) \qquad (6)$$

as demonstrated later in equations 9, 13, and 15. It is found convenient to incorporate this term in the equations in the following manner. First, E_M in the auxiliary equations for e_A, e_B, and e_C (equation 33, Appendix I) is redefined as

$$E_M = E_M' - 3Z_A K_1 (I_{PL} + i_T) \qquad (7)$$

in which E_M' is the equivalent source peak voltage and reduces to the same value for E_M for the single-rectifier case. Equations 1 and 2 are then unchanged in form. However, the impedance operator Z_A in equation 7 when expanded introduces a term containing the derivative of the d-c branch current on the right-hand side of equation 5. In the computer routine, equation 5 for the rate of change of i_T (d-c branch) was evaluated first, after collecting the derivative terms from both sides of the equation. This value was then used in evaluating the remaining rates of change of current.

The varying value of E_T, based on a complex expression for the voltage of the secondary bus given later, presents a problem if an attempt is made to include it in the equations already derived. Since its value generally varies slowly, however, a value calculated from the results at the last preceding time was used in making each new projection.

The computer routine devised for this problem has provision for including the effect of an additional impedance in the faulted branch. This was included to permit the calculation of the effect of breaker arc voltage. However, it may be readily adapted to indicate instead the effect of any nonlinear term such as tube arc drop.

Constants of the Mathematical Model

Equations 3 and 7 are written in terms of six distinct phase impedances Z_1, Z_2, Z_3, Z_4, Z_A, and Z_L, and two distinct system constants K_1 and K_2. These are defined in the "Nomenclature" section and their evaluation is discussed in detail in this section. The relationship of these constants and parameters to the physical system is made more evident by a separate consideration of the following several parts of the complete system:

1. The high-voltage a-c system.

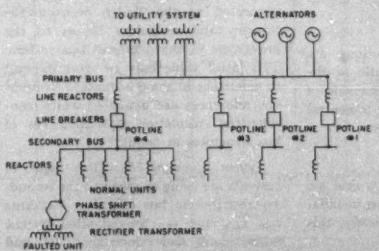
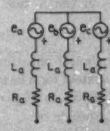


Fig. 6 (left). Single-line diagram of power system

Fig. 7 (below). Multiple sources reduced to single equivalent source



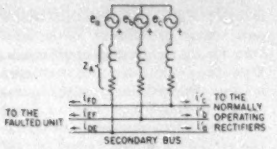


Fig. 8. Line currents at secondary distribution bus

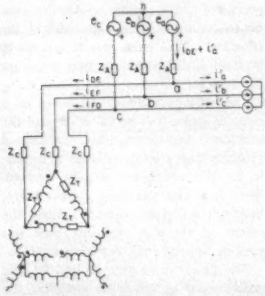


Fig. 9. Reduced high-voltage system with paralleling rectifiers replaced by equivalent current sources

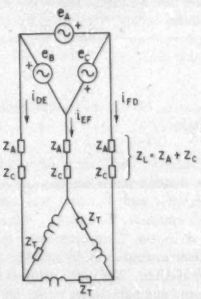


Fig. 10. Simplified circuit for high-voltage system with delta-connected source voltages

2. The low-voltage a-c system.
3. The interphase transformer.
4. The d-c system.

All of the following relations are derived assuming that the ratio of the line-to-line voltage on the delta side of the transformer to the line-to-neutral voltage on the Y side is unity.

High-Voltage A-C System

A single-line diagram of a typical a-c supply system for a large rectifier installation is shown in Fig. 6. Basically, this system consists of:

1. A primary distribution bus supplied by electric power company tie lines and on-the-site generators.

2. A system of overhead feeder busses connected to the primary distribution bus and each supplying a secondary distribution bus to which a number of rectifier transformers are connected.

3. Other auxiliary equipment such as circuit breakers, line reactors, phase-shift transformers, and regulating transformers.

Three impedances (Z_A, Z_L, and Z_1) and the two system constants (K_1 and K_2) will be fully defined upon referral of the high-voltage a-c system to the low-voltage side of the transformer.

Impedance Z_A. To begin, all the lines or generators feeding the primary distribution bus are reduced to an equivalent 3-phase source. For instance, the utility system lines and the alternators of Fig. 6 can be reduced by well-known techniques to the equivalent source of Fig. 7.

In this analysis the assumption is made that only one potline is energized. The effect of adding potline feeders or other load to the primary distribution bus is to cause the voltage to drop on this bus due to the impedance drop associated with the equivalent source impedance. This effect can be equivalently accounted for by using a smaller value of source voltage. If only a single potline feeder is considered, the equivalent circuit of Fig. 8 is an accurate representation of the actual circuit to the secondary distribution bus.

Note that

$$Z_A = Z_a + Z_b \tag{8}$$

Impedance Z_a (Fig. 5) is the equivalent impedance of the a-c system to the primary bus. Determination of this value is a routine procedure consisting of paralleling the equivalent impedance of the utility's supply system, as determined from knowledge of the system available short-circuit current and R/X ratio, and the impedances of the generators.

Impedance Z_b is obtained by adding the impedances of all the equipment connected in the feeder between the primary and secondary busses to the impedance of the overhead bus system. This latter value may be more difficult to determine in some instances. Numerous references are available which consider the calculation of inductances of bus structures in detail.[8-9]

System Constant K_1. Two sets of currents are being supplied to the secondary distribution bus (Fig. 8). Currents i_{DE}, i_{EF}, and i_{FD} are supplied to the faulted unit, and currents i_a', i_b', and i_c' are supplied to the normal rectifier units connected to this bus. This latter set is important in the analysis since these currents are responsible for the d-c feedback into the faulted unit. The assumption is made that the potline current and the d-c feedback are reflected into the a-c system as pure sinusoids in phase with the equivalent source voltages, or

$$\begin{aligned} i_a' &= K_1\sqrt{3}\ (\text{potline current} + \text{d-c feedback}) \sin(\omega t + \phi_a) \\ i_a' &= K_1\sqrt{3}\,(I_{PL} + i_T)\sin(\omega t + \phi_a) \\ i_b' &= K_1\sqrt{3}\,(I_{PL} + i_T)\sin(\omega t + \phi_b) \\ i_c' &= K_1\sqrt{3}\,(I_{PL} + i_T)\sin(\omega t + \phi_c) \end{aligned} \tag{9}$$

Upon inspection of these equations, it is evident that K_1 is the ratio of the maximum value of a-c line current to the magnitude of the total direct current or, under steady-state conditions,

$$K_1 = \frac{(I_{ac})M}{\sqrt{3}\,(I_{PL})} \tag{10}$$

Constant K_1 may be evaluated from current measurements on an existing system. If this is not possible or practical a value of 0.53 may be used as an approximation. The analysis on which this figure is based is given in Appendix II.

Transformer Secondary Voltages e_D, e_E, e_F. Having evaluated K_1, equation 9 defines values for i_a', i_b', and i_c', and the a-c system feeding the normally operating rectifiers may be replaced by current sources as illustrated in Fig. 9. This in turn can be further reduced by redefining the source voltages.

The line-to-neutral voltages at points a, b, and c on the secondary bus (Fig. 9) may be written

$$\begin{aligned} e_{an} &= e_a - Z_A(i_{DE} + i_a') = e_a - Z_A i_a' - Z_A i_{DE} \\ e_{bn} &= e_b - Z_A(i_{EF} + i_b') = e_b - Z_A i_b' - Z_A i_{EF} \\ e_{cn} &= e_c - Z_A(i_{FD} + i_c') = e_c - Z_A i_c' - Z_A i_{FD} \end{aligned} \tag{11}$$

where

$$\begin{aligned} e_a &= \frac{E_M'}{\sqrt{3}} \sin(\omega t + \phi_a) \\ e_b &= \frac{E_M'}{\sqrt{3}} \sin(\omega t + \phi_b) \\ e_c &= \frac{E_M'}{\sqrt{3}} \sin(\omega t + \phi_c) \end{aligned} \tag{12}$$

Now, if e_a', e_b', and e_c' are defined as

$$\begin{aligned} e_a' &= e_a - Z_A i_a' \\ e_b' &= e_b - Z_A i_b' \\ e_c' &= e_c - Z_A i_c' \end{aligned} \tag{13}$$

equation 11 can be rewritten as

$$e_{an}=e_a'-Z_A i_{DE}$$
$$e_{bn}=e_b'-Z_A i_{EF}$$
$$e_{cn}=e_c'-Z_A i_{FD} \quad (14)$$

Finally, if the Y-connected source consisting of e_a', e_b', and e_c' is converted to a delta-connected source, the circuit of Fig. 10 results, where

$$e_A=e_c'-e_a'=E_M \sin(\omega t+\phi_1)$$
$$e_B=e_a'-e_b'=E_M \sin(\omega t+\phi_2)$$
$$e_C=e_b'-e_c'=E_M \sin(\omega t+\phi_3) \quad (15)$$

where E_M is defined by equation 7.

The circuit of Fig. 10 is easily referred to the low-voltage side by the same procedure illustrated earlier for the single-rectifier unit. The detailed procedure for this latter case is included in Appendix I. The resulting equations are identical with equation 32 except for the redefinition of voltage indicated in equation 7.

Impedance Z_L. From Fig. 10, using equation 8 for Z_A,

$$Z_L=Z_a+Z_b+Z_c \quad (16)$$

where Z_c is obtained similarly to Z_b by adding the impedances of the equipment in the feeder between the secondary bus and the rectifier transformer bushings.

Impedance Z_1. Substituting equation 16 into equation 31 and noting that Z_T of Fig. 10 is equal to Z_d of Fig. 5 yields the defining relation for Z_1.

$$Z_1=3(Z_a+Z_b+Z_c)+Z_d \quad (17)$$

Low-Voltage A-C System

Impedance Z_2. The component parts of this system include:

1. The conductors connecting the low-voltage side of the rectifier transformer to the anode breakers.
2. The anode breakers.
3. The conductors connecting the anode breakers to the anode of the rectifier tubes.
4. The mercury-arc rectifier tubes.

The resistance and inductance of the conductors are best obtained by applying fundamental concepts. The resistance of the tube was assumed to be a constant obtained by averaging the resistance of the tube obtained as the ratio of tube voltage to tube current for a number of load conditions. The impedance of the breaker when the contacts are closed is negligible. On those occasions when the current-limiting action of the breaker was being investigated, the voltage characteristic of the breaker was used in preference to an equivalent impedance. This voltage was added as a voltage drop in the faulted branch after the breaker contacts parted.

Impedance Z_2, then, is simply a summation of the various impedances in the circuit between the low-voltage side of the rectifier transformer and the unit cathode bus.

Interphase Transformer

Impedance Z_3. The hysteresis loop of the material used in the core of the interphase is approximated by a linearized loop as shown in Fig. 3(A). Three different ranges of operation can be recognized, namely, Regions *I*, *II*, and *III*. In each individual region the slope of the hysteresis loop is constant, signifying a constant inductance. Each range is defined in terms of the net magnetizing currents at its boundaries. To determine the inductance in any given range use is made of the following expressions for inductance.

$$L=\frac{N^2 A 10^{-8}}{l}\frac{dB}{dH} \quad (18)$$

where

N = number of turns
A = area of core, square inches
l = length of path in core, inches
dB/dH = slope of **B-H** characteristic, (lines/square inch)/(ampere-turns/inch)

Mutual inductance M_{ip1} is equal to L_{ip1} under the assumption that the coefficient of coupling is unity and under the additional assumption that both coils have an equal number of turns. Both of these assumptions are made in this analysis. If these assumptions are not valid, M_{ip1} is calculated as

$$M_{ip1}=K\sqrt{L_{ip1}L_{ip2}} \quad (19)$$

The equivalent circuit for the interphase transformer is taken to be that of Fig. 3(B).

Impedance Z_3 (equal to Z_s in Fig. 5) is defined as

$$Z_3=R_{ip1}+(L_{ip1}+M_{ip1})\frac{d}{dt} \quad (20)$$

or

$$Z_3=R_{ip1}+2L_{ip1}\frac{d}{dt} \quad (21)$$

D-C System

Impedance Z_4. The equivalent circuit of each of the normal rectifier units is identical with the circuit of Fig. 2. The normal units are operating essentially in steady state and accordingly, except during commutation, only one phase on each Y is conducting at a time. This situation remains in effect even under the condition that these units are supplying feedback to the faulted unit. Thus, in agreement with previous reasoning and test results, the impedance of each of these units is taken to be,[6]

$$Z_q=\frac{Z_1+Z_2+Z_3}{2}+R_{COM}+Z_f+Z_h \quad (22)$$

where R_{COM} is an equivalent resistance

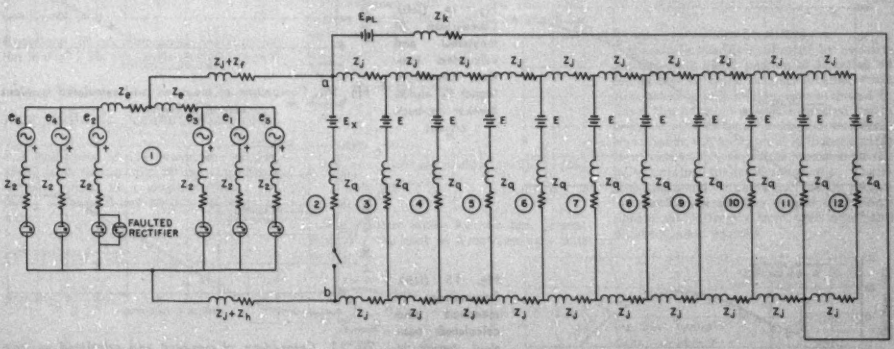


Fig. 11. Equivalent circuit of d-c system

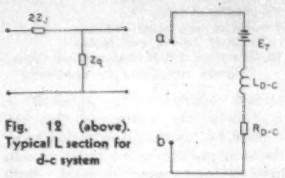


Fig. 12 (above). Typical L section for d-c system

Fig. 13 (right). D-c circuit equivalent for branch 7 of rectifier circuit

required to account for the steady-state commutation drop. On this basis the over-all circuit for the total system can be equivalently represented as shown in Fig. 11.

Since in the usual system the potline inductance is relatively large compared with the other system components, the paralleling effect of this branch on the remainder of the circuit is negligible. Accordingly, in reducing the d-c system to a simple equivalent, the potline is assumed to be a constant-current device with an infinite internal impedance. The circuit that remains is composed of a number of L sections, as indicated in Fig. 12, connected in tandem. It is possible to reduce this network between *a* and *b*, Fig. 11, to a simple equivalent as indicated in Fig. 13. A procedure is indicated in Appendix III. The equivalent of the d-c branch impedance is combined with the vertical branch of the interphase equivalent circuit, Fig. 3(B), to yield

$$Z_4 = R_{dc} + R_f + R_h + 2R_j + (L_{dc} + L_f + L_h + 2L_j)\frac{d}{dt} \quad (23)$$

System Constant K_2. Concerning the second system constant it is pertinent to note that the normal rectifier units which supply the d-c system during the transient period are excited by the secondary bus voltage. Any change in this voltage will result in a corresponding change in the direct voltage. As was mentioned earlier, the equivalent direct voltage is accordingly made proportional to this bus voltage. System constant K_2 is the constant of proportionality, defined by

$$K_2 = \frac{E_7}{(E_{ac})_M} = \frac{E_7}{\sqrt{2}\,E_{ac}} \quad (24)$$

which may be evaluated for an existing system from direct measurements of the d-c and secondary a-c bus voltages under full load.

If K_2 must be calculated it is remembered that, for an ideal rectifier,

$$E_{do} = \frac{3}{\pi}\sqrt{\frac{3}{2}}\,E_{ac} = 1.17E_{ac} \quad (25)$$

and it is found that to a good approximation

$$E_7 = E_{do} - I_{PL}R_{dc} \quad (26)$$

Then, it follows that

$$K_2 = \frac{E_7}{\sqrt{2}E_{ac}} = 0.827 - \frac{I_{PL}R_{dc}}{(E_{ac})_M} \quad (27)$$

Under steady-state conditions, the line-to-neutral voltages of equation 11 will yield

$$(E_{ac})_M = E_M' - 3Z_A K_1 I_{PL} \quad (28)$$

which may be used in the evaluation of K_2.

In the course of the calculations for arc-back current, $(E_{ac})_M$ must be evaluated and then multiplied by K_2 to find E_7. However, equation 28 no longer applies. Due to the unbalanced currents feeding the faulted rectifier, the phase voltages at the secondary bus are no longer balanced symmetrical values. Nevertheless, any reasonable method of averaging the phase voltages should provide a satisfactory value for the mean secondary bus voltage E_{ac} which may then be used to obtain E_7. The averaging method used by the authors is described in Appendix IV.

Study of an Actual System

A study was made of an actual installation. The procedure followed in the analysis was as indicated in this paper. The results of this study are compared with results of tests performed on the same system. This comparison is presented in Figs. 14 through 17 where the observed and calculated values of i_3, i_4, i_5, and i_{dc} are compared. A complete description of this system, namely the Rockdale Plant of the Aluminum Company of America, located in Texas, and a description of the tests are presented in other papers.[1,7]

Appendix I. Branch Voltages of Single-Rectifier Model

The high-voltage side of the system shown in Fig. 1 is readily transformed to the equivalent circuit of Fig. 18 in which the transformer has been replaced by an equivalent 1-to-1 transformer with potential terms e_D, e_E, and e_F and lumped leakage impedances Z_T; the per-unit value of the line impedances is unchanged, and the generator has been replaced by an equivalent delta-connected generator. The currents in the transformer windings are written

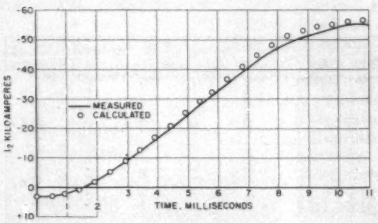


Fig. 14 (left). Comparison of measured and calculated transient current in branch 2, anode breaker arc-back current

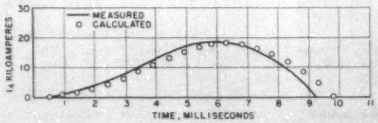


Fig. 15 (left). Comparison of measured and calculated transient current in branch 4

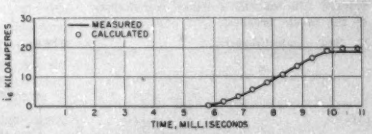


Fig. 16. Comparison of measured and calculated transient current in branch 6

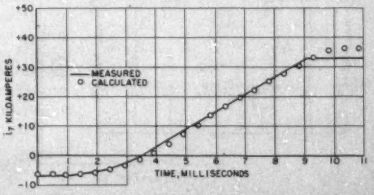


Fig. 17. Comparison of measured and calculated transient current in d-c branch: cathode breaker current

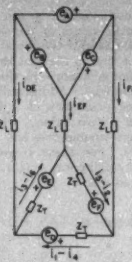


Fig. 18 (left). Simplified circuit for single-rectifier high-voltage system

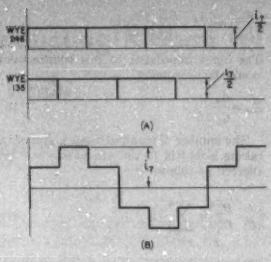


Fig. 19 (right). Simplified representation of d-c and a-c waveforms

A—D-c output
B—A-c demand

directly from Fig. 1 with the sign conventions of Fig. 2. Applying Kirchhoff's laws at the nodes and around the loops of the circuit of Fig. 18, the following equations may be written

$$\begin{aligned} e_A - Z_L(i_{PD} - i_{DE}) - e_D - Z_T(i_1 - i_4) &= 0 \\ e_B - Z_L(i_{DE} - i_{EF}) - e_E - Z_T(i_2 - i_4) &= 0 \\ e_C - Z_L(i_{EF} - i_{PD}) - e_F - Z_T(i_3 - i_4) &= 0 \end{aligned} \tag{29}$$

$$\begin{aligned} i_{PD} - i_{DE} &= (i_1 - i_4) - (i_4 - i_2) - (i_3 - i_4) + (i_1 - i_4) \\ &= 3(i_1 - i_4) - [(i_1 - i_4) + (i_4 - i_2) + (i_2 - i_4)] \\ &= 3(i_1 - i_4) + i_{1p1} \end{aligned} \tag{30}$$

$$\begin{aligned} i_{DE} - i_{EF} &= 3(i_2 - i_4) + i_{1p1} \\ i_{EF} - i_{PD} &= 3(i_3 - i_4) + i_{1p1} \end{aligned}$$

Let

$$Z_1 = Z_T + 3Z_L \tag{31}$$

Then, substituting equation 30 into equation 29 and solving for e_D, e_E, and e_F gives

$$\begin{aligned} e_D &= e_1 = -e_4 = e_A - Z_1(i_1 - i_4) - Z_L i_{1p1} \\ e_E &= e_2 = -e_4 = e_B - Z_1(i_2 - i_4) - Z_L i_{1p1} \\ e_F &= e_3 = -e_4 = e_C - Z_1(i_3 - i_4) - Z_L i_{1p1} \end{aligned} \tag{32}$$

where the source voltages are taken to be

$$\begin{aligned} e_A &= E_M \sin(\omega t + \phi_1) \\ e_B &= E_M \sin(\omega t + \phi_2) \\ e_C &= E_M \sin(\omega t + \phi_3) \end{aligned} \tag{33}$$

Equations 32 are the phase voltages for use in Fig. 2 for the single-rectifier system.

Appendix II. Computation of K_1

A half-cycle of the waveform of Fig. 19(B) is shown in Fig. 20 in which $2a$ is the d-c output of a single rectifier. Defining the waveform as $F(\theta)$, it may be written

$$\begin{aligned} F(\theta) &= a; \quad 0 \le \theta < \frac{1}{3}\pi \\ F(\theta) &= 2a; \quad \frac{1}{3}\pi \le \theta < \frac{2}{3}\pi \\ F(\theta) &= a; \quad \frac{2}{3}\pi \le \theta < \pi \end{aligned} \tag{34}$$

This function $F(\theta)$ may be written as

$$F(\theta) = B_1 \sin\theta + B_2 \sin 2\theta + B_3 \sin 3\theta + \ldots + B_n \sin n\theta + \ldots \tag{35}$$

where

$$B_n = \frac{2}{\pi}\int_0^{\pi} F(\theta) \sin n\theta d\theta \tag{36}$$

Substituting $F(\theta)$ from equation 34 into equation 36, gives

$$\begin{aligned} B_n &= \frac{2a}{\pi}\left(\int_0^{\pi/3} \sin n\theta d\theta + 2\int_{\pi/3}^{2\pi/3} \sin n\theta d\theta + \int_{2\pi/3}^{\pi} \sin n\theta d\theta\right) \\ &= \frac{2a}{\pi n}\left(1 + \cos\frac{1}{3}n\pi - \cos\frac{2}{3}n\pi - \cos n\pi\right) \end{aligned} \tag{37}$$

whence

$$\begin{aligned} B_1 &= \frac{6}{\pi}a \\ B_2 &= B_3 = B_4 = 0 \\ B_5 &= \frac{6}{5\pi}a \\ B_6 &= 0 \\ B_7 &= \frac{6}{7\pi}a \end{aligned} \tag{38}$$

For rectifiers with a uniform phase-shift angle α, the resulting current waveform for each harmonic is[8]

$$\begin{aligned} &\sum_{m=0}^{m=p-} B_n \sin n(\theta + m\alpha) \\ &= B_n \frac{\sin\frac{1}{2}p n\alpha}{\sin\frac{1}{2}n\alpha} \sin n\left[\theta + \frac{1}{2}(p-1)\alpha\right] \end{aligned} \tag{39}$$

The Fourier series for the line current supplying the bank of p rectifiers may thus be written

$$i_{ac} = \sum_{n=1}^{n=\infty} B_n \frac{\sin\frac{1}{2}p n\alpha}{\sin\frac{1}{2}n\alpha} \sin n\left[\theta + \frac{1}{2}(p-1)\alpha\right] \tag{40}$$

$$= \frac{6a}{\pi}\left\{\frac{\sin\frac{1}{2}p\alpha}{\sin\frac{1}{2}\alpha}\sin\left[\theta + \frac{1}{2}(p-1)\alpha\right] + \frac{1}{5}\frac{\sin\frac{5}{2}p\alpha}{\sin\frac{5}{2}\alpha}\sin 5\left[\theta + \frac{1}{2}(p-1)\alpha\right] + \frac{1}{7}\frac{\sin\frac{7}{2}p\alpha}{\sin\frac{7}{2}\alpha}\sin 7\left[\theta + \frac{1}{2}(p-1)\alpha\right] + \ldots\right\} \tag{41}$$

Considering only the fundamental term, the amplitude is

$$(i_{ac})_M = 2ap\left(\frac{3}{\pi}\right)\frac{\sin\frac{1}{2}p\alpha}{p\sin\frac{1}{2}\alpha} \tag{42}$$

If the phase-shift transformers are picked to give uniform spacing, as is commonly the case,

$$\alpha = \frac{\frac{1}{3}\pi}{p} \tag{43}$$

and

$$\frac{\sin\frac{1}{2}p\alpha}{p\sin\frac{1}{2}\alpha} = \frac{\sin\frac{\pi}{6}}{p\sin\frac{\pi}{6p}} \approx \frac{6}{\pi}\sin\frac{\pi}{6} = \frac{3}{\pi} \tag{44}$$

the latter being a good approximation for a small value of α or a large value of p. Since $2a$ is the d-c output of one rectifier, $2ap$ would be the potline current corresponding to $(i_{ac})_M$. Thus equations 42 and 44 give

$$(i_{ac})_M \approx \frac{9}{\pi^2} I_{PL} \tag{45}$$

and

$$K_1 = \frac{(i_{ac})_M}{\sqrt{3}I_{PL}} = \frac{3\sqrt{3}}{\pi^2} = 0.526 \tag{46}$$

for a large number of rectifiers.

It may be readily confirmed by evaluating the exact (left) portion of equation 44 for as few as three or four rectifiers that this value is still within better than 1%. Since the harmonics were neglected and a much simplified branch current waveform was used in the analysis, it is inappropriate to attach a high degree of precision to the value obtained for small values of p. Therefore the value in equation 46 is recommended for use with any number of rectifiers phased off uniformly from each other and dividing load equally.

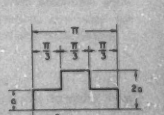


Fig. 20. Typical delta, double-Y a-c waveform

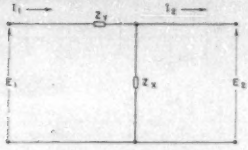


Fig. 21. Typical d-c circuit element

Appendix III. Evaluating the Equivalent Impedance of the D-C System

Input Impedance of a Circuit

Consider Fig. 21. The input voltage and current of this circuit can be related to the output voltage and current, in terms of $A\ B\ C\ D$ constants as follows

$$E_1 = AE_2 + BI_2 \tag{47}$$

$$I_1 = CE_2 + DI_2 \tag{48}$$

or, in matrix form

$$\begin{Vmatrix} E_1 \\ I_1 \end{Vmatrix} = \begin{Vmatrix} A & B \\ C & D \end{Vmatrix} \cdot \begin{Vmatrix} E_2 \\ I_2 \end{Vmatrix} \tag{49}$$

In general the input impedance would be

$$Z_{in} = \frac{E_1}{I_1} = \frac{AE_2 + BI_2}{CE_2 + DI_2} \tag{50}$$

which, if $I_2 = 0$, reduces to

$$Z_{in} = \frac{A}{C} \tag{51}$$

Application to Actual System

This relationship also applies in the general case. For instance, if n L sections of the type shown in Fig. 21 were cascaded, the relationship between the input and output quantities would be, in matrix form,

$$\begin{Vmatrix} E_{in} \\ I_{in} \end{Vmatrix} = \begin{Vmatrix} A & B \\ C & D \end{Vmatrix} \begin{Vmatrix} A & B \\ C & D \end{Vmatrix} \cdots \begin{Vmatrix} A & B \\ C & D \end{Vmatrix} \begin{Vmatrix} E_{out} \\ I_{out} \end{Vmatrix} \tag{52}$$

or

$$\begin{Vmatrix} E_{in} \\ I_{in} \end{Vmatrix} = \begin{Vmatrix} A & B \\ C & D \end{Vmatrix}^n \begin{Vmatrix} E_{out} \\ I_{out} \end{Vmatrix} \tag{53}$$

If

$$\begin{Vmatrix} A & B \\ C & D \end{Vmatrix}^n = \begin{Vmatrix} A_s & B_s \\ C_s & D_s \end{Vmatrix} \tag{54}$$

then

$$\begin{Vmatrix} E_{in} \\ I_{in} \end{Vmatrix} = \begin{Vmatrix} A_s & B_s \\ C_s & D_s \end{Vmatrix} \begin{Vmatrix} E_{out} \\ I_{out} \end{Vmatrix} \tag{55}$$

The input impedance to this combination would be (assuming $I_{out} = 0$)

$$Z_{in} = \frac{A_s}{C_s} \tag{56}$$

The number of manipulations involved in raising a matrix to the nth power may be effected as follows[9, 10]

$$\begin{Vmatrix} A & B \\ C & D \end{Vmatrix}^n = \sum_{r=1}^{2} (\lambda_r)^n \frac{\prod_{s \neq r} (\lambda_s I - [a])}{\prod_{s \neq r} (\lambda_s - \lambda_r)} \tag{57}$$

where the λ's are defined by the determinant

$$\begin{vmatrix} A - \lambda & B \\ C & D - \lambda \end{vmatrix} = 0 \tag{58}$$

The $A\ B\ C\ D$ constants of the 4-terminal network of Fig. 21 are

$$\begin{Vmatrix} A & B \\ C & D \end{Vmatrix} = \begin{Vmatrix} 1 + \frac{Z_y}{Z_x} & Z_y \\ \frac{1}{Z_x} & 1 \end{Vmatrix} \tag{59}$$

Alternate Method

The required impedance may alternately be obtained by assuming a unit voltage applied to the input terminal, calculating the response current, and dividing the voltage by this current. To determine the response current, a set of simultaneous equations may be written and solved readily on an electronic computer.

This method has the merit of permitting the actual impedance of the potline to be used in the calculations, removing the need for assuming infinite impedance. However, the large ratio of impedances normally found between the potline and the rectifiers results in little difference between the two procedures.

Appendix IV. Determination of the Mean Voltage on the Secondary Bus Under Transient Conditions

The individual phase voltages at the secondary bus are given by equation 11. Perhaps the most obvious method for obtaining an arbitrary "average" amplitude would be to apply an equation of the form

$$\frac{1}{\sqrt{3}}(E_{ac})_M = \frac{1}{3}\left[\frac{e_{an}}{\sin(\omega t + \phi_a)} + \frac{e_{bn}}{\sin(\omega t + \phi_b)} + \frac{e_{cn}}{\sin(\omega t + \phi_c)}\right] \tag{60}$$

However, this is not practical to use in computation. At small values of the sine terms in the denominators considerable distortion takes place due to the fact that the numerators are not in phase with their respective denominators. Additionally, there is the recurring possibility of a denominator taking on a zero value. Division by zero is always troublesome and in this case intolerable. For these reasons another method of averaging was sought.

The relation selected was

$$\frac{1}{\sqrt{3}}(E_{ac})_M = \frac{2}{3}[e_{an}\sin(\omega t + \phi_a) + e_{bn}\sin(\omega t + \phi_b) + e_{cn}\sin(\omega t + \phi_c)] \tag{61}$$

Computationally, this equation presents no difficulties. It has the effect of giving greatest weight to the phases nearest their peak amplitude assuming that the phase displacement between the weighting factors and the actual voltage waves is not large. Substitution of balanced symmetrical voltages in phase with the sinusoidal weighting factors will verify that equation 61 yields a value equal to their common amplitudes, a necessary requirement for any equation selected.

References

1. Comparison of Calculated and Measured Arc-Back Current in Large Power Rectifier Systems, J. Teno, C. H. Titus, R. N. Wagner. *AIEE Transactions*, vol. 77, pt. II, 1958 (*Paper no. 58-223*).

2. Numerical Mathematical Analysis (book), J. B. Scarborough. The Johns Hopkins Press, Baltimore, Md., 1950, pp. 235–40.

3. Formulas for the Calculation of the Inductance of Linear Conductors of Structural Shape, Thomas James Higgins. *AIEE Transactions*, vol. 62, Feb. 1943, pp. 53–58.

4. Formulas for the Inductance of Rectangular Tubular Conductors, T. J. Higgins. *Ibid.*, vol. 60, Dec. 1941, pp. 1046–50.

5. Reactance Values for Rectangular Conductors, H. B. Dwight. *Electric Journal*, London, England, vol. 16, 1919, pp. 255–56.

6. Rectifier Fault Currents, C. C. Herskind, H. L. Kellogg. *AIEE Transactions*, vol. 64, Mar. 1945, pp. 145–50.

7. Anode Breaker Testing With a High-Capacity Rectifier System, L. J. Harris, T. J. Scully, V. N. Stewart. *AIEE Transactions*, vol. 77, pt. II, 1958 (*Paper no. 58-224*).

8. Tables of Integrals and Other Mathematical Data (book), H. B. Dwight. The Macmillan Company, New York, N. Y., revised edition, 1947, eq. 420.3.

9. Mathematics of Modern Engineering, Vol. II (book), E. G. Keller. John Wiley & Sons, Inc., New York, N. Y., 1942, p. 237.

10. Transmission Lines and Filter Networks (book), John J. Karakash. The Macmillan Company, 1950, pp. 405–07.

Temperature of Conducting Ribbon Magnet Coils

JOHN F. H. DOUGLAS
FELLOW AIEE

THIS PAPER applies the theory of heat flow to finding the steady-state temperature in excess of the surrounding iron. The ribbon coils may be either copper or aluminum, and may be either two or four in number. The magnets considered are large magnets of circular cylindrical shape such as lifting magnets.

A cross section of a typical magnet is shown in Fig. 1, showing the iron, coils, and part of the insulation. The dashed line is the axis of revolution. Coil dimensions include: axial ribbon width h, inner radius r_a, and outer radius r_b. The insulation shown comprises: 1. mica plates, 2. mica cylinder, and 3. thermosetting substance. Not shown is the turn insulation, which may be thin asbestos paper, or (in the case of aluminum) aluminum oxide formed by electrolysis. To compute average temperature, all dimensions, properties of materials, and watts must be known.

It is assumed that steel temperature can be estimated. It is assumed that the steel is isothermal at its average temperature. It is assumed that the temperature of the ribbon conductor does not vary in the axial direction.

This paper discusses the following topics: 1. statement of the differential equation, 2. solution of the differential equation, 3. illustrative example, and 4. conclusions. The proofs are found in Appendix I and the nomenclature in Appendix II.

The Differential Equation

If reference is made to Fig. 2, the computed value of temperature for a typical magnet, it will be seen that a relation between temperature and the radius is needed. This relation is found in the differential equation 1.

$$d^2\theta/dr^2+(1/r)(d\theta/dr)-(g_p/h\mu_r)(\theta) = -(W'/h\mu_r) \quad (1)$$

Paper 58-809, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22-27, 1958. Manuscript submitted July 5, 1957; made available for printing April 10, 1958.

JOHN F. H. DOUGLAS is with Marquette University, Milwaukee, Wis.

In this equation θ is the temperature, in degrees centigrade (C), above the surrounding iron. The quantity r is the radius in inches. The quantity g_p is the specific thermal conductivity in watts per C, for the mica plate, and for one square inch. The quantity h is the width of the ribbon in inches. The quantity μ_r is the thermal conductivity of the ribbon coil in a radial direction, in watts per C for an inch cube of the coil including metal and insulation. The watts per square inch of mica plate is denoted by the symbol W'.

Solution to the Differential Equation

There are two solutions to the differential equation. The first is a functional solution permitting plotting of curves such as in Fig. 2. This solution is in Appendix I. The more practical solution gives an expression for the average temperature in terms of five thermal constants This equation is

$$\theta_{av}=\theta_p-\theta_a f_a(r_a)-\theta_b f_b(r_b) \quad (2)$$

The terms θ_p, θ_a, and θ_b are shown in Fig. 2.

In equation 2 the term θ_p is given by equation 3. The terms θ_a and θ_b are given by equations 4 through 6, and the factors f_a and f_b are computed by equations 7 through 9. Further information is given under "Thermal Constants" on the computation of specific conductance from conductivity using equations 10 and 11.

The chief term in equation 2 is that denoted by θ_p. It may be found by considering all the heat to flow in an axial direction through the mica plates. If W' denotes the watts per square inch for the boundary mica plate, then the value of θ_p for two coils is given by equation 3(A). For four coils, the average value of the principal terms is given by equation 3(B).

$$\theta_p=(W'/g_p) \quad (3A)$$

$$\theta_p=1.25(W'/g_p) \quad (3B)$$

With four coils equation 3(A) applies to the outer coils, but the inner coils are hotter since half of the watts must flow through the inner mica layer. By averaging their results, equation 6 is obtained.

The boundary values of the minor terms in equation 2 are the constants θ_a and θ_b. They are determined from the condition that the heat flowing through the coil to the boundary must equal that flowing through the boundary layer to the iron. They are computed by equations 4 and 5.

$$\theta_a=\left[\frac{\theta_p}{1+(\mu_r C/g_a)F_a(Cr_a)}\right] \quad (4)$$

$$\theta_b=\left[\frac{\theta_p}{1+(\mu_r C/g_b)F_b(Cr_b)}\right] \quad (5)$$

Note that the factors F_a and F_b are given in Table I and are functions of C times the boundary radius.

In equations 4 and 5 the term g_a is the thermal conductance of one square inch of the mica ring, while g_b is the thermal conductance of one square inch of the outer insulation, both in watts per C. The constant C in equations 4 and 5 is an abbreviation for the value given by equation 6.

$$C=(g_p/h\mu_r)^{0.5} \quad (6)$$

The remaining factors in equation 1 are the terms f_a and f_b. Their values are

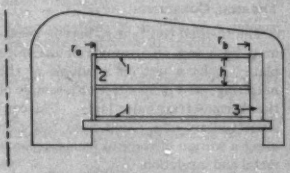


Fig. 1. Cross section of magnet with two aluminum ribbon coils, center line shown. Coil dimensions are h, r_a, and r_b. Insulation is marked as follows: mica, 1 and 2; thermosetting substance, 3

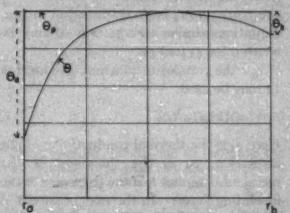


Fig. 2. Curve plotted from functional solution of differential equation, computed for coil made of anodized aluminum ribbon

θ = temperature
θ_p = principal component
θ_a, θ_b = minor components due to radial heat flow
r_a, r_b = boundary radii, shown as abscissas

Table I. Factors for Finding Boundary Temperatures

Cr	F_a(Cr)	F_b(Cr)
3.0	1.15	0.81
3.5	1.13	0.84
4.0	1.12	0.86
4.5	1.105	0.88
5.0	1.095	0.89
5.5	1.09	0.905
6.0	1.08	0.91
7.0	1.07	0.925
8.0	1.06	0.935
9.0	1.055	0.94
10.0	1.05	0.95
11.0	1.045	0.955
12.0	1.04	0.96

computed from equations 8 and 9. These equations have a term V denoting a unit volume, defined as the volume of the lower coil in one radian, and given by equation 7. Thus,

$$V=0.5(h)(r_b^2-r_a^2) \quad (7)$$

$$f_a=(h/C^2V)(Cr_a+0.46) \quad (8)$$

$$f_b=(h/C^2V)(Cr_b-0.46) \quad (9)$$

Thus, equations 3 through 9 supply all the information needed to use equation 2 to compute average temperature.

Thermal Constants

The thermal constant g_p may be measured by a test of a sample of the mica plate; g_a by a test on a sample of the mica ring; and g_b by a test on a sample of the thermosetting outer layer. The constant (μ_r) may be found from test also using a section of the coil including both metal and insulation.

These constants may also be computed if the thermal conductivities of the several materials and the thickness of the insulating layer are known. For example

$$g_p=\mu_p/t_p \quad (10)$$

where μ_p is the thermal conductivity of the plate and t_p is its thickness in inches. Equations similar to equation 10 may be written for g_a and g_b.

For the composite constant μ_r equation 11 may be used.

$$\mu_r=(\mu_i)[1+(t_m/t_i)] \quad (11)$$

Here μ_i is the thermal conductivity of the insulation, t_m is the thickness of the metal ribbon, and t_i is the thickness of the insulating layer. In the case of anodized aluminum ribbon, the experimental method is preferred.

Illustrative Example

Fig. 2 was computed for a magnet with the radii r_a of 8 inches and r_b of 24 inches. The magnet was assumed to have two aluminum coils of ribbon 3.1 inches wide, with insulation formed by anodizing. The mica plates were assumed to be 0.2 inch thick, the mica cylinder 0.1 inch thick, and the thermosetting layer 1 inch thick. The watts per square inch of mica plate was assumed to be 2.5.

The thermal conductivity of mica was assumed to be 0.02 watts per C for an inch cube. The thermal conductivity of the thermosetting layer was assumed to be 0.008 watt per C for an inch cube. The radial conductivity for the coil composite was assumed to be 0.2 watt per C for an inch cube of coil.

Using equation 10 the thermal constant g_p was computed to be 0.1, the constant g_a as 0.2, and g_b as 0.008. The constant C given by equation 6 was computed as 0.4. Next the quantity θ_p was computed by equation 3, giving 25 C. Quantities θ_a and θ_b were computed by equations 4 and 5 as 17 C and 3 C respectively.

The factors f_a and f_b were computed by equations 7 through 9 and the results were put into equation 2 with the result that the average temperature was 22.8 C above the surrounding iron.

Conclusions

1. A short method of engineering accuracy for finding the average temperature of ribbon coils is presented.

2. For the same space factor two ribbon coils are cooler than four ribbon coils.

3. Consistent with durability thinner turn insulation results in cooler coils due to improved space factor and better thermal properties.

4. The illustrative problem assuming anodized aluminum ribbon appears to offer promise. Tests on the durability of this material should be watched with interest.

Appendix I. Proofs

The several sections of the proofs are discussed under the subheadings: 1. the differential equation, 2. the particular solution, 3. complementary solutions, 4. boundary equations, and 5. average temperature.

The Differential Equation for Steady-State Temperature

The coil element considered is shown in Fig. 3. This comprises a section of the coil of height (h), subtending one radian, and between radius r and radius $(r+\delta r)$. If δW_1 represents heat produced in the element (by Joule's law), δW_2 the heat escaping axially, and (δW_3) that escaping radially, then for a steady-state temperature,

$$\delta W_1-\delta W_2-\delta W_3=0 \quad (12)$$

If W' represents the heat generated per

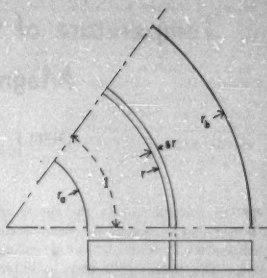


Fig. 3. Coil element used in proofs, element is between radii r and $(r+\delta r)$ and subtends one radian

square inch of mica plate then, in the area $(r)(\delta r)$,

$$\delta W_1=W'(r)(\delta r) \quad (13)$$

To find the axial flow of heat, temperature, conductance, and area are multiplied. Thus

$$\delta W_2=(\theta)(g_p)(r)(\delta r) \quad (14)$$

Two steps are needed to find the third term δW_3. First the radial heat flow through the element must be found, and then the amount by which this varies between radius r and radius $(r+\delta r)$ must be found.

The radial flow of heat through the coil element is the product of the radial conductivity μ_r, the area rh, and the temperature gradient. Thus,

$$W_r=-(\mu_r)(hr)(d\theta/dr) \quad (15)$$

The variation in this quantity or the excess of the outer over the inner radius is

$$\delta W_3=(d'W_r/dr)(\delta r)=-(h\mu_r)(\delta r)(d/dr)\times(rd\theta/dr) \quad (16)$$

$$=-(h\mu_r)(\delta r)[r(d^2\theta/dr^2)+(d\theta/dr)]$$

When equations 13 through 15 are substituted into equation 12, and the equation is divided by $(h\delta r)$, equation 17 is obtained.

$$d^2\theta/dr^2+(1/r)(d\theta/dr)-(g_p/h\mu_r)(\theta)=-(W'/h\mu_r)$$

or

$$d^2\theta/dr^2+(1/r)(d\theta/dr)-C^2\theta=-(W'/h\mu_r) \quad (17)$$

Like all second-order equations there are three terms, i.e., the particular solution and two complementary solutions. The particular solution in this case is the part that is constant and the two complementary solutions adjust for the radial boundaries. In general the particular solution eliminates the second member of the equation. The differential equation 17 is known as Bessel's equation of the modified type.

The Particular Solution of the Temperature Equation

If in equation 17 the temperature θ is assumed to be constant, and its derivatives

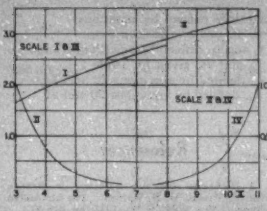


Fig. 4. Curves used for graphical integration to find average temperature. Curves I and II are for inner boundary and III and IV are for outer boundary, curves II and IV are from exponential terms, while I and III are from Bessel functions, in asymptotic form

zero, then the solution is equation 3(A). This part of the solution could be found by simple reasoning, independent of the differential equation. It represents the temperature drop in the mica plate if the heat is assumed to flow only in an axial direction. With four ribbon coils, the inner coils had a θ_p that is 50% larger and the average is given by equation 3(B).

The Two Complementary Solutions for Coil Temperature

The complementary solutions for Bessel's equation are the solutions of equation 17 when the constant term θ_p is eliminated. Thus the complementary solutions are found from equation 18.

$$d^2\theta/dr^2+(1/r)(d\theta/dr)-C^2\theta=0 \qquad (18)$$

The functions which satisfy equation 18 are called Bessel functions of the modified type. For the larger boundary r_b, the function is said to be of the zero order and first kind and is denoted by the function $I_0(Cr)$. For the inner boundary r_a, the function is said to be of the zero order and second kind and is denoted by the function $K_0(Cr)$. These functions are tabulated in Watson's text.[1]

If it is desired to plot a curve of temperature versus radius such as in Fig. 2, then equation 19, which supplies the arbitrary constants needed to match the boundary conditions, is used.

$$\theta=\theta_p-\theta_a\left[\frac{K_0(Cr)}{K_0(Cr_a)}\right]-\left[\frac{I_0(Cr)}{I_0(Cr_b)}\right] \qquad (19)$$

If, however, it is desired to compute average temperature, then it is more convenient to use the asymptotic forms given in Watson's text.[2] Thus in the second term

$$K_0(Cr)/K_0(Cr_a)=e^{-C(r-r_a)}(r_a/r)^{0.5}\times[S_K(Cr)/S_K(Cr_a)] \qquad (20)$$

and if the abbreviation x is used for Cr, the S functions are given by

$$S_K(x)=1-(1/8x)+(9/128x^2)-(75/1,024x^3)+\ldots \qquad (21)$$

In the third term of equation 19 the bracket is given by equation 22.

$$I_0(Cr)/I_0(Cr_b)=e^{C(r-r_b)}(r_b/r)^{0.5}\times[S_I(Cr)/S_I(Cr_b)] \qquad (22)$$

where again if the abbreviation (x) is used in place of (Cr), the S_I function is given by the series

$$S_I(x)=1+(1/8x)+(9/128x^2)+(75/1,024x^3)+\ldots \qquad (23)$$

The equations in this section are not needed for the body of the paper, they are used in a later section to evaluate equations for average temperature.

If it is desired to visualize the manner in which these Bessel functions vary, then equations 20 through 23 are of help. The most important factor is the exponential term. Second in importance is the square-root term. Of minor importance is the function S_K or the function S_I.

Boundary Equations

The boundary values θ_a and θ_b are determined by equating the heat flowing through the coil to the boundary to that flowing through the boundary. The superior boundary is discussed in detail first. In the proof, the following notation is used for the derivative of $I_n(x)$.

$$(d/dx)[I_0(x)]=I_1(x) \qquad (24)$$

The heat leaving the outer boundary is proportional to g_b and to $(\theta_p-\theta_b)$. The heat approaching the boundary is proportional to μ_r and to the temperature gradient. Thus,

$$(g_b)(\theta_p-\theta_b)=-(\mu_r)(d\theta/dr)=-[\mu_r\theta_b/I_0(Cr_b)]CI_1(Cr) \qquad (25)$$

If Cr is replaced by x, for brevity, then the asymptotic expression 26 is useful

$$F_b=I_1(x)/I_0(x)=1-(1/2x)-(1/8x^2)-(1/8x^3)-(25/128x^4)-\ldots \qquad (26)$$

This factor in Table I was computed from equation 26 using Cr_b in place of x. With this factor, equation 25 becomes equation 27.

Then the factor F_b is used in place of the derivative and equation 25 can be more briefly written

$$g_b(\theta_p-\theta_b)=\theta_b\mu_r CF_b(Cr_b) \qquad (27)$$

When this equation is solved for θ_b, equation 5 results.

For the inner boundary an exactly similar procedure may be used, resulting in equation 7, where the factor F_a is given by equation 28.

$$F_a(x)=F_a(Cr)=K_1(x)/K_0(x)=1+(1/2x)-(1/8x^2)+(1/8x^3)-\ldots \qquad (28)$$

Note that the functions $I_1(x)$ and $K_1(x)$ are Bessel functions of the first order, of the modified type, and of the first and second kinds. They are tabulated in reference 1, which may be used if more precise values are needed than are given in Table I.

Calculation of Average Temperature of the Coils

The average temperature is found by an integration. The first task is to find the integral. The integral does not lend itself to analytical procedure. The second task is to outline steps which make graphical integration simpler. The integration covers

Table II. Factors Used in Finding Average Temperature

x	$\sqrt{x}S_K(x)$	$\sqrt{x}S_I(x)$
3.0	1.65	1.82
3.5	1.81	1.945
4.0	1.94	2.075
4.5	2.06	2.185
5.0	2.18	2.30
5.5	2.29	2.405
6.0	2.40	2.51
7.0	2.60	2.70
8.0	2.77	2.88
9.0	2.96	3.045
10.0	3.12	3.20
11.0	3.30	3.36
12.0	3.43	3.49

unit volume V defined by equation 9 as the volume of one radian of a single coil.

If R_1 is the resistance of a cubic inch at a selected base temperature (taken as that of the iron), if α is the temperature coefficient of resistance, and if (δR) is the increase of resistance, then

$$\delta R=(R_1\alpha)\int_{r_a}^{r_b}h(\theta)(dV)=(R_1\alpha)\int_{r_a}^{r_b}h(\theta)(r)(dr) \qquad (29)$$

If θ is taken as the average temperature, then equation 29 becomes

$$\delta R=(R_1\alpha\theta_{av})(V) \qquad (30)$$

When equations 29 and 30 are combined, the desired equation 31 results.

$$\theta_{av}=(1/V)(h)\int_{r_a}^{r_b}(\theta)(r)(dr) \qquad (31)$$

Or since it is preferable to change the variable r to (x/C) and (dr) to (dx/C), a convenient expression is

$$\theta_{av}=(h/VC^2)\int_{x_a}^{x_b}(\theta)(x)(dx) \qquad (32)$$

Equation 19 for temperature θ has three terms, and therefore equation 1 must have three. If θ is made θ_p in equation 31 then the first term of equation 1 is as shown. The procedure in graphically integrating equation 32 is shown in detail for the second term in equation 19 using equations 20 and 21 to obtain the expression

$$\theta_a f_a(h/VC^2)(\theta_a)\left[\frac{\sqrt{x_a}}{S_K(x_a)}\right]\times\int_{x_a}^{x_b}e^{-(x-x_a)}\sqrt{x}\,S_K(x)dx \qquad (33)$$

Equation 33 directs one to compute the exponential term and the product of the two remaining terms in the integrand. These are shown in curves *I* and *II* in Fig. 4, for the case of $x_a=3$. Curve *I* was plotted from Table II. Equation 33 directs one to multiply the two curves together, find the area under this product, and then multiply by the bracket preceeding the integral sign. This was done for a number of values of x_a. The several values thus found were plotted against x_a and found to be a straight line with a probable error of 1%. The result is contained in equation 8. A similar procedure

was used for the third term in equation 19, with the result contained in equation 9.

Appendix II. Nomenclature

g = specific thermal conductance, watts per square inch per C
h = width of ribbon, inches
r = radius, inches
R_i = resistance of an inch cube of the coil, at base temperature, ohms
t = thickness of insulation, inches
W' = watts per square inch of mica plate
V = unit volume of coil, defined by equation 7, cubic inches
x = a variable replacing Cr
α = temperature coefficient of resistance
θ = temperature above the surrounding iron, C
μ_r = thermal conductivity for an inch cube, watts per C

Subscripts:
a = inner boundary
b = outer boundary
i = turn insulation
p = plate insulation or particular solution
r = radial direction

Quantities defined by equations:
C, defined by equation 6
f_a and f_b, defined by equations 8 and 9
F_a and F_b, given by Table I
I_b, defined by equations 22 and 23
K_b, defined by equations 20 and 21

References

1. A Treatise on the Theory of Bessel Function (book), G. N. Watson. The Macmillan Company, New York, N. Y., 1948, pp. 698–713.
2. *Ibid.*, pp. 202–03, equations 1, 2, arts. 7, 23.

Equivalent Circuits for 3-Phase 4-Wire Induction Motors Operating on Nonsymmetrical Systems

J. T. MITCHELL
ASSOCIATE MEMBER AIEE

CALCULATION METHODS for predicting the performance of 3-phase 4-wire induction motors operating on nonsymmetrical systems are not generally known. Equivalent circuits and calculation procedures for polyphase motors with floating neutral have been available for some time.

Polyphase motors used to drive accessories on aircraft almost always have the neutral bonded to the airframe. Open circuits in one or more of the line leads caused by a loose connection or open circuit breaker are quite common. The motor must continue to operate and in many cases start under load with one line open. Cold starting at −55 degrees centigrade to −65 degrees centigrade is a customary requirement. To complicate the problem further is the fact that there is very little margin of safety in design. Motor size is greatly influenced by the torque characteristics of the load. The lighter and smaller motor that will meet the specified requirements is preferred even if it costs more than a larger motor. It is important for the designer to predict motor performance during adverse conditions such as variable frequency, variable voltage, extremely cold temperatures, high ambient temperatures, one line open, two lines open, unbalanced voltages, etc.

Nonsymmetrical operation reduces the torque capabilities and increases the watts loss for a given torque. Torque required to breakaway motor bearings and moving parts of the load at −55 degree centigrade can be quite high on gear boxes, hydraulic pumps, air compressors, and other oil or grease-lubricated devices. Speed–torque characteristics and temperature rise data of the motor for adverse conditions are invaluable for proper motor application.

The method of symmetrical components is a valuable tool for analyzing motor performance on unbalanced systems. General symmetrical component equivalent circuits for unbalanced systems leave a large gap to be bridged before they can be conveniently used by design engineers to calculate motor performance. The purpose of this paper is to extend these circuits to include motor circuits in their conventional form and establish the relationship between the constants required for the extended circuit and the conventional constants. In addition, the relationship between zero-sequence and third space harmonic fields is developed.

Paper 58-844, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing May 12, 1958.

J. T. Mitchell is with the Westinghouse Electric Corporation, Lima, Ohio.

The author wishes to acknowledge the valuable advice and guidance of H. J. Braun.

The contents of this paper represent a portion of the thesis submitted by the author to the University of Pittsburgh, Pittsburgh, Pa., in partial fulfillment of the requirements for the Master of Science degree.

Review of Fundamental Principles

Consider a 3-phase induction motor wound full pitch and connected to a source having balanced terminal voltages as shown in Fig. 1. The sum of line current I_a, I_b, I_c, must equal zero at any instant of time. A forward revolving field (direction of rotation assumed forward) will be produced in the motor by the current resulting from the fundamental component of applied voltage. Time harmonics of the fundamental will also set up revolving magnetic fields. Some of these fields rotate forward and some backward.

Direction of rotation is considered forward when the harmonic has the same phase sequence as the fundamental and backward when the phase sequence is opposite that of the fundamental. The mathematical equations for the line-to-neutral voltages are given in Appendix I. Phasor representation of the phase voltages is given in Fig. 2.

It will be observed that the 5th harmonic rotates backward, the 7th forward, the 11th backward, and the 13th forward. By omitting tripleen harmonics, the remaining harmonics occur in pairs. It will be found that the lower order of a pair rotates backward and the higher order rotates forward. Thus, the 11th time harmonic produces a magnetic field in the motor which revolves backward at 11 times fundamen-

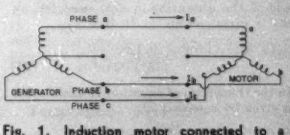


Fig. 1. Induction motor connected to a balanced source

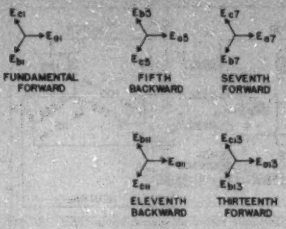


Fig. 2. Vector representation of time harmonics

tal synchronous speed. Each of the time harmonics shown in Fig. 2 forms a balanced system of 3-phase voltages and resulting line currents will be balanced.

The third harmonic and its multiples are all in time phase (of zero sequence) and have no return path for current. Consequently, there can be no magnetic field set up in the motor by them (it is possible under certain conditions for a generator to produce a balanced 3-phase set of triple frequency voltages 120 deg (degrees) out of phase, but these are not true third harmonics of the fundamental and will not be considered here).

If the neutral is connected, the return path is complete for third time harmonic currents. These currents produce a magnetic field which is stationary in space and pulsates at three times fundamental frequency. As shown in Fig. 3, this magnetic field will have three times as many poles as the fundamental. Following the analysis by the double revolving field theory for a single-phase motor, the magnetic field produced by third harmonic currents can be considered as two component fields of half-magnitude revolving in opposite directions. The component fields will revolve at fundamental synchronous speed since revolutions per minute = 120 (frequency) /poles.

Any harmonic which is a multiple of the third sets up a magnetic field which also has three times the number of fundamental poles. However, the component fields of the ninth harmonic would revolve at three times fundamental synchronous speed, etc. The magnetic field having three times the number of fundamental poles is produced because the windings of the motor carry currents which are in time phase. It is interesting to note that only the third harmonic and multiples of the third set up an independent field and do not produce any fundamental flux. All other time harmonics produce fundamental poles and space harmonic poles of the fundamental. Analysis of zero-sequence operation is the same as for the third harmonic in time except that zero-sequence currents are of fundamental frequency.

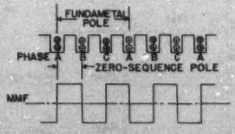


Fig. 3. Zero-sequence mmf (either fundamental or third time harmonic)

Nomenclature

a = vector operator = $\epsilon^{j\ 120\ \text{deg}}$
C = series conductors per phase
E_a = volts on phase a
E_{a0} = zero-sequence component of voltage on phase a
E_{a1} = positive-sequence component of voltage on phase a
E_{a2} = negative-sequence component of voltage on phase a
E_{a5} = fifth time harmonic voltage on phase a
E_1 = fundamental voltage
E_b = voltage on phase b
E_c = voltage on phase c
E_{g1} = positive-sequence generated voltage
f = frequency
$f(x)$ = function of x
I_a = current in phase a
I_b = current in phase b
I_c = current in phase c
I_{g0} = zero-sequence component of generator current
I_{g1} = positive-sequence component of generator current
I_{g2} = negative-sequence component of generator current
I_{m0} = current in magnetizing branch of zero-sequence circuit
I_z = current in load impedance
j = operator to rotate vector 90 deg
K_w = winding factor for single-phase motor = $KcKd$
K_{w0} = zero-sequence winding factor
K_{w1} = positive-sequence winding factor = $KcKd$
K_{w3} = winding factor for third space harmonic = K_{w0}
m = number of phases
N_c = conductors per slot
p = number of fundamental poles
q = slots per phase per pole
R_m = resistance included in magnetizing branch of single-phase motor to simulate core loss
R_{m1} = resistance included in positive-sequence circuit to simulate core loss
R_r = rotor resistance for single-phase motor
R_{r1} = rotor resistance of positive-sequence motor
R_{r3} = rotor resistance of third space harmonic motor
R_s = stator resistance of single-phase motor
R_{s1} = stator resistance for positive sequence
s = rotor slip with respect to positive-sequence field
t = unit of time
V_{d1} = voltage drop across positive-sequence rotor circuit
V_f = voltage across forward zero-sequence rotor circuit
ω = equal to $2\pi f$
X_m = magnetizing reactance for single-phase motor
X_r = rotor bar leakage reactance
X_{r1} = rotor leakage reactance for positive sequence
X_s = stator leakage reactance for single-phase motor
X_{s1} = stator leakage reactance for positive sequence
Z_0 = zero-sequence motor impedance
Z_1 = positive-sequence motor impedance
Z_2 = negative-sequence motor impedance
Z_{g0} = zero-sequence generator impedance
Z_{g1} = positive-sequence generator impedance
Z_{g2} = negative-sequence generator impedance
$\propto$ = symbol for "proportional to"
τ = pole pitch

Nature of Sequence Fields

Assuming that the motor as shown in Fig. 4 is supplied with unbalanced voltages, the basic equivalent circuits can be established. Only fundamental frequency is considered.

Assume:

$$E_a = 120/0\ \text{deg} \quad E_b = 115/240\ \text{deg} \quad E_c = 100/120\ \text{deg}$$

By the method of symmetrical components:

Zero-sequence voltage:

$$E_{a0} = (1/3)(E_a + E_b + E_c)$$
$$= (1/3)(120 + 115/240\ \text{deg} + 100/120\ \text{deg})$$
$$= (1/3)(120 - 57.5 - j99.6 - 50 + j86.6)$$
$$= (1/3)(12.5 - j13)$$
$$= 6/-46.2\ \text{deg}$$

Positive-sequence voltage:

$$E_{a1} = (1/3)(E_a + aE_b + a^2E_c)$$
$$= (1/3)(120 + 115 + 100)$$
$$= 111.7/0\ \text{deg}$$

Negative-sequence voltage:

$$E_{a2} = (1/3)(E_a + a^2E_b + aE_c)$$
$$= (1/3)(120 + 115/120\ \text{deg} + 100/240\ \text{deg})$$
$$= (1/3)(120 - 57.5 + j99.6 - 50 - j86.6)$$
$$= 6/46.2\ \text{deg}$$

Since the voltage from which the

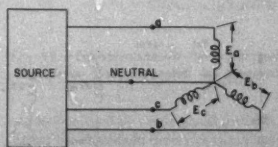


Fig. 4. Motor supplied with unbalanced voltages

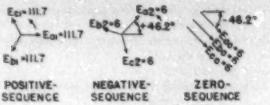


Fig. 5. Vector representation of sequence voltages

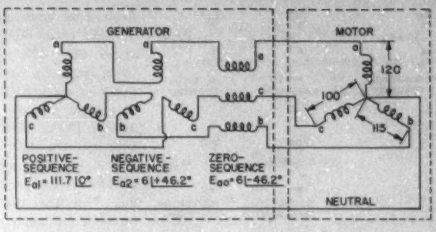


Fig. 6. Equivalent circuit for one condition of unbalance

components were evaluated is fundamental frequency, the component voltages are also of fundamental frequency. Vector representation of the sequence voltages is shown in Fig. 5. The motor is effectively supplied from three series-connected generators as shown in Fig. 6. Generator impedance is considered to be zero for this illustration.

The principle of superposition can be used provided the motor can be analyzed as a linear circuit. Actually a motor is not a linear device, since the resistance and reactance depend to some extent upon saturation and skin effect, which vary with load. However, approximations can usually be made and the motor reduced to a linear equivalent circuit.

Positive-sequence voltages applied to the terminals of the motor will result in balanced positive-sequence line currents and a magnetic field of constant magnitude rotating in the positive direction at fundamental synchronous speed. The well-known equivalent circuit as shown in Fig. 7 applies.

When negative-sequence voltages only are applied to the terminals, balanced line currents of negative sequence are produced and the resulting magnetic field is of constant magnitude but rotates in the negative direction (opposite the positive-sequence rotation) at fundamentals ynchronous speed. The equivalent circuit is the same as Fig. 7. However, if the shaft is revolved forward by external or by positive-sequence power, the equivalent circuit for negative sequence is shown in Fig. 8.

Zero-sequence voltages are all in time phase and do not form a balanced system of polyphase voltages as do the positive- and negative-sequence voltages. When zero-sequence voltages only are applied to the motor terminals, line currents will be of equal magnitude and in time phase. Their only return path is the neutral connection to the star point. The zero-sequence current in the neutral will be equal to the arithmetic sum of the zero-sequence line currents or three times the current in one line. Zero-sequence currents establish a magnetic field stationary in space and pulsating at fundamental frequency. This field will have three times as many poles as produced by positive or negative-sequence currents. There is a definite similarity between the action of zero-sequence currents and third time harmonic currents. The nature of the magnetic field produced in the motor is the same except that the zero-sequence currents are at fundamental frequency. By resolving the zero-sequence field into two fields of equal magnitude (each equal to one half the magnitude of the pulsating field) and rotating in opposite directions at one third of the fundamental synchronous speed, the equivalent circuit for zero-sequence is arrived at as shown in Fig. 9.

It should be remembered that S is the per-unit slip of the rotor with respect to the positive-sequence field.

A clear distinction should be made between the nature of the field established by zero-sequence currents and that produced by positive- (or negative) sequence currents. Positive-and/or negative-sequence currents produce no third space harmonic (or tripleen harmonic) magnetomative force (mmf) in a balanced winding. Zero-sequence currents produce only tripleen space harmonic mmf. Winding factors (k_{w0} = pitch factor times distribution factor) for the third space harmonic produced by zero-sequence are given in Table I.

If the line and phase voltages of the motor are known, the sequence components can be evaluated and applied to their respective circuits independently, provided the circuits are assumed to be linear. Thus, for the example of Table I, performance calculations would be made with 117.7/0 deg volts applied to circuit in Fig. 7, $\overline{6/46.2}$ deg applied to circuit Fig. 8, and $\overline{6/-46.2}$ deg applied to the circuit in Fig. 9. The resultant torque is equal to the algebraic sum of the component torques.

$$T_{net} = T_{pos} + T_{neg} + T_{f\ zero} + T_{b\ zero} \quad (1)$$

The zero-sequence forward torque, $T_{f\ zero}$, becomes negative for $s < 2/3$. The negative-sequence torque T_{neg}, and the zero-sequence backward torque $T_{b\ zero}$, are negative for all forward speeds. Line currents can be evaluated from their sequence components.

$$I_a = I_{a0} + I_{a1} + I_{a2} \quad (2)$$

$$I_b = I_{a0} + a^2 I_{a1} + a I_{a2} \quad (3)$$

$$I_c = I_{a0} + a I_{a1} + a^2 I_{a2} \quad (4)$$

Total power input is equal to the sum of the power input to each circuit. Power factors can be taken as:

$$\text{Power factor} = \frac{\text{watts}}{3EI} = \frac{\text{watts}}{3(\text{average phase voltage}) \times (\text{average phase current})} = \frac{3(\text{total power input})}{(|E_a| + |E_b| + |E_c|)(|I_a| + |I_b| + |I_c|)} \quad (5)$$

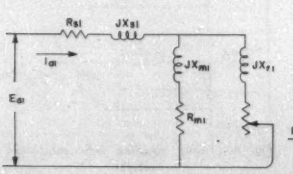
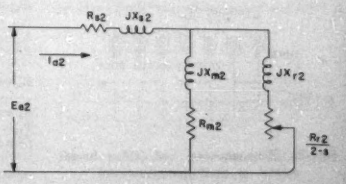


Fig. 7 (left). Equivalent circuit for motor supplied with positive sequence

Fig. 8 (right). Equivalent circuit for motor supplied with negative-sequence voltage and driven forward



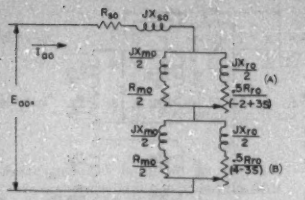


Fig. 9. Equivalent circuit for motor supplied with zero-sequence voltage

A—Forward field
B—Backward field

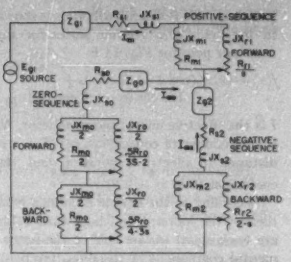


Fig. 10. Complete equivalent circuit for 3-phase motor operating on two phases

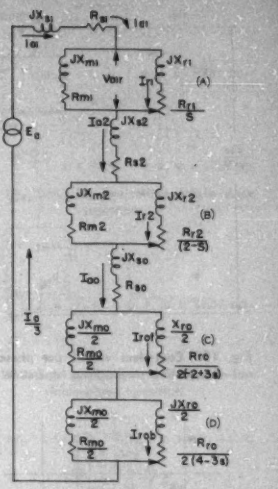


Fig. 11. Equivalent circuit per phase for line-to-neutral single-phase operation of a 3-phase motor

A—Positive-sequence forward
B—Negative-sequence backward
C—Zero-sequence forward
D—Zero-sequence backward

When the motor current is a small portion of the total load current supplied by an unbalanced system, the sequence voltages may be calculated from measured voltages and applied to the respective equivalent motor circuits. Calculations for three or four load points at approximately 62% slip should be made to determine the minimum pull-up torque. Locked rotor torque and full load losses should also be determined to obtain adequate application information.

Equivalent Circuits

Equivalent circuits for some of the common nonsymmetrical conditions can be obtained by connecting the motor circuits to the source per the general circuits given Wagner and Evans.[1] The equivalent circuit for one line missing (special case of infinite series impedance in one line) is given in Fig. 10. It must be remembered that the sequence voltages are evaluated from the motor terminal voltages. Line-to-neutral voltage on the open phase is very low with the rotor locked. The most important load points are full load, locked rotor, and minimum pull-up torque at approximately 62% slip.

The equivalent circuit for line-to-neutral single-phase operation is given in Fig. 11. Since there is no net starting torque for this condition, usually full load and maximum torque are the only points of interest. Line and alternator impedances may be lumped with the stator impedance for easier solution.

One of the most common causes of unbalanced voltages on systems is heavy single-phase loads. Quite often a large portion of the total load is composed of a few single-phase loads that are switched on and off as required. Large motors may reduce the unbalance considerably, since the motor current will be larger in the phase having the highest voltage. The basic symmetrical component circuit of an alternator supplying a 3-phase load in parallel with a line-to-ground fault through an impedance represents this condition. Equivalent circuits for the motor can be included in this network as shown in Fig. 12.

Line-to-line single-phase operation has been analyzed by several authors[1,3,4] and the equivalent circuit is shown in Fig. 13. Absence of zero-sequence or third space harmonic fields are of particular interest and will be discussed later.

Table I. Zero-Sequence Winding Factors

q	Pitch	kwo
1	1	1.0000
1	1	0.0000
2	2/3	0.0000
2	1	0.7071
2	5/6	0.5000
2	2/3	0.0000
2	1/2	0.5000
3	1/3	0.7071
3	1	0.6667
3	8/9	0.5773
3	7/9	0.3333
3	2/3	0.0000
3	1/3	0.6667
4	1	0.6533
4	11/12	0.6035
4	5/6	0.4619
4	3/4	0.2500
4	2/3	0.0000
5	1	0.6472
5	14/15	0.6155
5	4/5	0.4000
5	11/15	0.2000
5	2/3	0.0000
6	1	0.6439
6	17/18	0.6220
6	8/9	0.5577
6	5/6	0.4553
6	7/9	0.3220
6	13/18	0.1667

Relation of Sequence Fields

It is interesting to note that the equivalent circuit for the condition of single phase line-to-neutral operation includes a zero-sequence network in addition to a forward and backward field per the common double revolving field theory of a single-phase motor. Actually, the motor

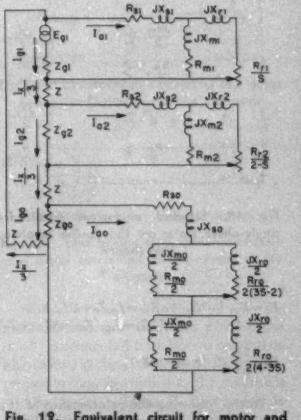


Fig. 12. Equivalent circuit for motor and single-phase load connected to generator

can be analyzed by single-phase methods. Either the cross-field theory or the double revolving field theory can be used. Terminal voltage on phases *b* and *c* can be found by the following relations:

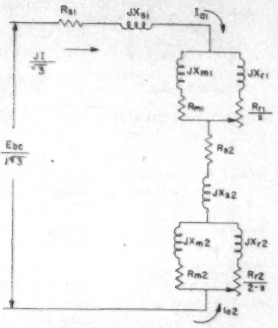


Fig. 13. Equivalent circuit per phase for line-to-line single-phase operation

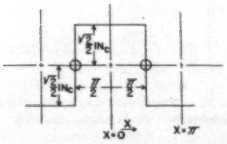


Fig. 14. Mmf of a full pitch coil

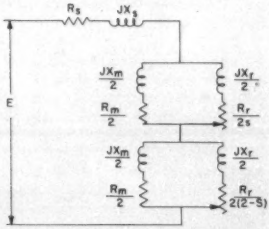


Fig. 15. Standard equivalent circuit for single-phase motor per the double revolving field theory

$$E_b = E_{a0} + a^2E_{a1} + aE_{a2} = I_{a0}Z_0 + a^2I_{a1}Z_1 + aI_{a2}Z_1 = I_{a1}(Z_0 + a^2Z_1 + aZ_2)$$

$$E_c = \frac{I_a}{3}(Z_0 + aZ_1 + a^2Z_2)$$

In the following the relationship of the positive-sequence and negative-sequence fields to the forward and backward fields evaluated by the double revolving field theory will be investigated. Also, why the zero-sequence network appears in the equivalent circuit for a single-phase motor will be determined.

Assume the motor has one slot-per-pole per-phase and is wound full pitch as shown in Fig. 14. The Fourier series for the mmf per pole for Fig. 14 is

$$f(x) = 0.9IN_c(\cos x - 1/3\cos 3x + 1/5\cos 5x - 1/7\cos 7x + \ldots)$$

I is the effective curent per conductor, N_c is the number of conductors per slot, x is distance in electrical degrees, and the coefficient (0.9) is $\sqrt{2}\,2/\pi$.

According to the double revolving field theory, the forward and backward fields are each equal to one half of the fundamental component of the pulsating field. The standard equivalent circuit is shown in Fig. 15.

Fundamental mmf:

$$\text{Forward field mmf/pole} = 0.45IN_c \quad (6)$$

$$\text{Backward field mmf/pole} = 0.45IN_c \quad (7)$$

Third space harmonic mmf:

$$\text{Forward field mmf/harmonic pole} = \frac{0.45IN_c}{3}$$

$$\text{Backward field mmf/harmonic pole} = \frac{0.45IN_c}{3} \quad (8)$$

Symmetrical component analysis:

Component currents have been evaluated as:[1]

$$I_{a1} = I_{a2} = I_{a0} = I_{a/3} \quad (9)$$

The mmf per pole for a polyphase machine is $m/2$ times the mmf of one phase.[2] Therefore:

$$\text{Positive-sequence } \frac{\text{mmf}}{\text{pole}} = \frac{m}{2}(0.9)N\frac{I_a}{3} = 0.45I_aN_c \quad (10)$$

and

$$\text{Negative-sequence } \frac{\text{mmf}}{\text{pole}} = \frac{m}{2}(0.9)\frac{N_aI_c}{3} = 0.45I_aN_c \quad (11)$$

Unlike the positive- and negative-sequence fields, the zero-sequence field is stationary in space and pulsating sinusoidally in time. The forward and backward zero-sequence fields are each equal to one half the pulsating component. Fig. 3 shows how zero-sequence currents in the three phases set up three times the number of fundamental poles.

This winding can be considered as a concentric winding having one full pitch coil per zero-sequence pole. The zero-sequence current is equal to $I_a/3$. Therefore, the peak mmf per zero-sequence pole is equal to $0.9\,N_aI_a/3$ consequently, the revolving component fields are as follows.

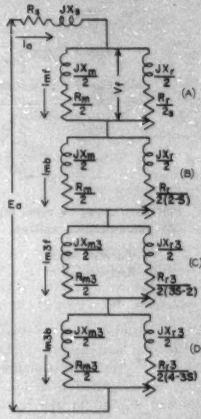


Fig. 16. Equivalent circuit for a single-phase motor with third space harmonic motor included

A—Forward field
B—Backward field
C—Third space harmonic, forward field
D—Third space harmonic, backward field

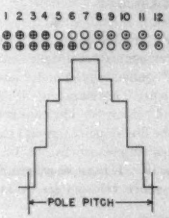


Fig. 17. Mmf of two phases carrying currents in time phase

Mmf per zero-sequence pole:

$$\text{Forward field} = 0.45\frac{N_cI_a}{3}$$

$$\text{Backward field} = 0.45\frac{N_aI_c}{3} \quad (12)$$

It will be observed that the positive-sequence field is identical to the forward revolving field and the negative-sequence field is identical to the backward revolving field of the common double revolving field theory for single-phase motors. Also, the zero-sequence field is identical to the third space harmonic field, which is usually neglected in single-phase motor analysis.

The equivalent circuit for a single-phase motor with the third space har-

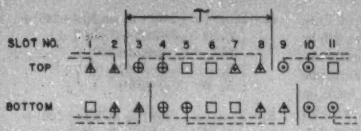



Fig. 18. Mmf for two phases carrying currents in time phase

monic included is shown in Fig. 16. Some authors have also included the fifth and seventh harmonic motors as a refinement. The equivalence of Fig. 11 and Fig. 16 is proved in Appendix II.

Equation 27 is rather interesting, since it has always been a problem to calculate accurately the stator leakage of a single-phase motor with concentric windings. Perhaps a good approach would be to replace the concentric windings with an equivalent distributed winding and calculate X_1 and X_{12} as 3-phase reactances. Normally, X_{11} and X_{12} are assumed to be equal under single-phase operation. It should be remembered that a two-thirds pitch does not eliminate $X_{\phi\phi}$. Although there will be no net zero-sequence air-gap flux with a two-thirds pitch, there will still be end-turn leakage and also some stray slot leakage because of the mmf's of the top and bottom coil sides do not cancel at all points.

Fig. 13 is quite similar to Fig. 15, but the constants are derived by different methods. Rotor resistance and leakage reactance constants are proportional to $m(CK_w)^2$. For the 3-phase constants for Fig. 13, C is the series conductors for one phase. However, for single-phase analysis, C is the series conductors for both phases or twice that of one phase.

It can be shown that the winding factor (K_w) for the single-phase analysis is $\sqrt{3}/2$ times the winding factor for the 3-phase analysis. As a numerical example, consider a 36-slot 4-pole 3-phase winding having a 7/9 pitch as shown in Fig. 17.

The winding factor for single-phase (line-to-line) analysis derived from Fig. 17 is 0.78108. K_{w1} for the 3-phase winding is 0.90191. These are in the ratio of $\sqrt{3}/2$. It may be stated in general terms for any given permeance that for the 3-phase motor:

$$\text{Reactance} = 3(CK_{w1})^2(2\pi f \text{ permeance})\ 10^{-8}$$

and for single-phase analysis $m=2$ and

$$\text{Reactance} = 2\left[(2C)\left(\frac{\sqrt{3}}{2}K_{w1}\right)\right]^2 (2\pi f \text{ permeance})\ 10^{-8}$$

$$= 6(CK_{w1})^2\ (2\pi f \text{ permeance})\ 10^{-8}$$

Since one half of the reactance is assigned to the forward field and one half to the backward field of the single-phase circuit, each field is given the value:

$$\text{Reactance per field} = 3(CK_{w1})^2(2\pi f)\ (\text{permeance})\ 10^{-8}$$

Constants for stator resistance for Fig. 13 is the sum of R_{11} and R_{12} or twice that of one phase. All that is necessary to convert Fig. 13 to Fig. 15 is to replace $E_{ac}/j\sqrt{3}$ by E_{ac} (the line-to-line voltage) and the sequence currents by the line current.

The symmetrical component analysis shows that the zero-sequence component does not exist for this connection. As would be expected, the third space harmonic is also absent regardless of the chording or the number of slots. Assume a winding having the distribution ($q=2$, and there is one slot short chord) as shown in Fig. 18.

This winding can be analyzed as two separate windings, one in the top of the slots and one in the bottom of the slots. It will be observed that coil sides of both the top and bottom windings contributing to each pole have an equivalent winding pitch of $2q$. (Refer to Fig. 19 and Fig. 20.) The top coil sides will be displaced from the bottom due to chording, but the pitch will remain $2q$ regardless of the chording. Note that a pitch of $2q$ is two thirds of the fundamental pole and is equivalent to 360 deg for the third harmonic. Another line of reasoning is that phases a and b are 60 deg apart in space and the mmf's of the two phases are equal in magnitude and same distribution. Therefore, any third space harmonic produced by phase a will be $3 \times 60 = 180$ deg out of space phase with the third harmonic produced by phase b.

Six possible distributions for the bottom coil sides are shown in Fig. 21. Active and idle coil sides are indicated by "1" and "0" respectively.

When these are analyzed as concentric windings it will also be found that the third space harmonic is zero.

It has been shown that the zero-sequence field and the third space harmonic field are identical. With the symmetrical component analysis of the motor, the zero-sequence field is automatically included when present. Normally, single phase calculations consider only the fundamental and neglect all space harmonics. On single-phase operation of a 3-phase

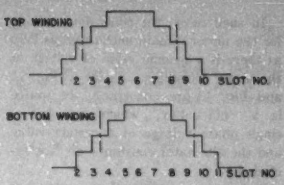


Fig. 19. Mmf for line-to-line single-phase operation

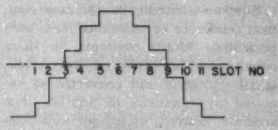


Fig. 20. Total mmf per pole for line-to-line single-phase operation

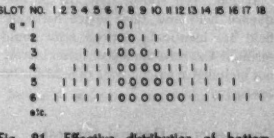


Fig. 21. Effective distribution of bottom coil sides of two phases carrying currents which are in time phase

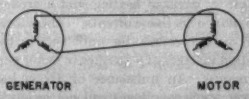


Fig. 22. Line-to-neutral operation of a motor connected line-to-line at the alternator

Fig. 23. Line-to-line single-phase operation of a motor connected line-to-neutral at the alternator

motor, it makes no difference whether the motor is connected from line-to-neutral or line-to-line at the alternator as long as the proper voltage is applied to the terminals of the motor. Thus, if the motor is rated 115 volts per phase, one line and neutral might well be connected to a 115-volt line-to-line alternator as shown in Fig. 22.

Likewise, single-phase line-to-line operation of a 3-phase motor can be obtained by connecting line-to-neutral at the alternator as shown in Fig. 23.

In such cases, the equivalent circuits for the motor remain unchanged as long as there is a specific value of voltage at the motor terminals. As far as Fig. 15 and Fig. 13 are concerned, the source is a "black box," which furnishes a single-phase voltage of a certain value and the alternator constants are not included.

Conclusions

Equivalent circuits for the most common conditions of unbalance have been presented. Motor constants for these circuits can be evaluated by conventional equations and converted as required by equations 13 and 26. For accurate analysis, close approximations must be made for saturated values of X_{m2} and X_{m0} and for skin effect on the rotor.

It has been shown that the zero-sequence field and third space harmonic field are identical. Serious error could result if the zero-sequence network were not included in calculations of unbalanced operation.

Comments of General Interest

Operation of a 3-phase 4-wire induction motor on an unbalanced system will result in increase heating and a greater unbalance in line currents. The motor will tend to reduce the unbalance in the system voltages at a sacrifice in motor performance. An unbalance of 8% or less in the applied voltages will not seriously affect speed–torque characteristics except possibly on a very marginal design.

With one line lead disconnected, torque capabilities are greatly reduced (in the order of 50%) and the I^2R losses may increase as much as 50% even on a conservatively rated rotor. The negative-sequence current produces a backward revolving field which results in a drag torque at all forward speeds. Zero-sequence currents produce a pulsating field which has three times as many poles as the positive-sequence field. It is essentially a single-phase field and can be analyzed by single-phase methods. An induction cusp will occur at +1/3 positive-sequence synchronous speed. The zero-sequence field, and the resulting cusp, can be eliminated by a winding pitch of 2/3, $1^1/_3$-, or by 120-deg phase belts (zigzag connection). However, the impedance of the motor to third time harmonic voltages will be greatly reduced and triple-frequency neutral current may be of sufficient magnitude to cause trouble with communications equipment. If the alternator has a winding pitch of 2/3-, $1^1/_3$-, or 120-deg phase belts, there will be no generated third time harmonic voltages. However, third time harmonic currents may be drawn by some loads on the system. The motor and alternator will furnish return paths for these currents.

Single-phase line-to-neutral and single-phase line-to-line operation are essentially the same except that the zero-sequence field is not produced with the single-phase line-to-line connection.

Under balanced conditions it is usually unimportant whether or not the neutral is connected. On some of the early models of alternators used on aircraft, the coil pitch was not two thirds and a third harmonic voltage appeared on the line. Most fractional pitch motors offer a relatively low impedance to third time harmonic currents when the neutral is connected. The third harmonic current in the neutral can be equal to the fundamental current in one line with only 5% third harmonic voltage at the alternator. A slight increase in stator copper loss would be observed under this condition. Usually the third harmonic current returning through the airframe was the most serious consideration. Elimination of the interference with electronic equipment required the use of reactors and occasionally the motor neutral was allowed to float.

Torque capabilities are only slightly reduced with 15% negative-sequence voltage. However, the motor losses are approximately doubled. Prohibitive temperatures may be reached on continuous-duty motors at high altitudes.

Unbalanced voltages are normally caused by large single-phase loads. The motor will draw more current from the phase that has the highest voltage and decrease the unbalance. For a particular condition of unbalance, the performance can be calculated by applying the proper positive and negative sequence voltages to Figs. 7 and 8 respectively. Calculation procedures are adequately described in the references.[1,3,4]

Appendix I

Mathematical expressions for the applied voltage are helpful in visualizing the difference between tripleen harmonics and other time harmonics. Assume that the applied voltage contains the 3rd, 5th, 7th, 9th, 11th, and 13th time harmonics.

Then:

$$E_a = E_1 \sin \omega t + E_3 \sin 3\omega t + E_5 \sin 5\omega t + E_7 \sin 7\omega t + E_9 \sin 9\omega t + E_{11} \sin 11\omega t + E_{13} \sin 13\omega t$$

$$E_b = E_1 \sin (\omega t + 240) + E_3 \sin 3(\omega t + 240) + E_5 \sin 5(\omega t + 240) + E_7 \sin 7(\omega t + 240) + E_9 \sin 9(\omega t + 240) + E_{11} \sin 11(\omega t + 240) + E_{13} \sin 13(\omega t + 240)$$

$$E_c = E_1 \sin (\omega t + 120) + E_3 \sin 3(\omega t + 120) + E_5 \sin 5(\omega t + 120) + E_7 \sin 7(\omega t + 120) + E_9 \sin 9(\omega t + 120) + E_{11} \sin 11(\omega t + 120) + E_{13} \sin 13(\omega t + 120)$$

After simplifying:

$$E_b = E_1 \sin (\omega t + 240) + E_3 \sin 3\omega t + E_5 \sin (5\omega t + 120) + E_7 \sin (7\omega t + 240) + E_9 \sin 9\omega t + E_{11} \sin (11\omega t + 120) + E_{13} \sin (13\omega t + 240)$$

$$E_c = E_1 \sin (\omega t + 120) + E_3 \sin 3\omega t + E_5 \sin (5\omega t + 240) + E_7 \sin (7\omega t + 120) + E_9 \sin 9\omega t + E_{11} \sin (11\omega t + 240) + E_{13} \sin (13\omega t + 120)$$

Phasor representation of the voltages is given in Fig. 2.

Appendix II

The per-phase equivalent circuit for line-to-neutral operation of a 3-phase motor is given in Fig. 11. This circuit is the exact equivalent of the circuit for a single-phase motor with third space harmonic included as shown in Fig. 16.

Constants for Fig. 11 are calculated in ohms per phase for the 3-phase machine. The numerical value of all positive and negative-sequence constants, except stator resistance, is proportional to the number of phases m.[2]

$$X_{m1} \propto m;\ X_{s1} \propto m;\ X_{r1} \propto m;\ R_{r1} \propto m$$

$$X_{m2} \propto m;\ X_{s2} \propto m;\ X_{r2} \propto m;\ R_{r2} \propto m$$

For single-phase motors, m is taken as two instead of unity. The constants are derived from the same permeances and rotor cage regardless of whether a 3-phase or a single-phase analysis is used.

If it is assumed that the reactance of the forward and backward field are equal as is usually done in single-phase analysis, then the ratio of the constants is easily established.

Thus,

$$X_{m1} = \frac{3}{2} X_m;\ X_{r1} = \frac{3}{2} X_r;\ R_{r1} = \frac{3}{2} R_r$$

$$X_{m2} = \frac{3}{2} X_m;\ X_{r2} = \frac{3}{2} X_r;\ R_{r1} = \frac{3}{2} R_r \qquad (13)$$

The voltage drop (V_f) across the secondary branch which represents the forward revolving field in Fig. 16 is

$$V_f = \frac{I_a \left[\left(\frac{R_r}{s} + jX_r\right)\right](R_m + jX_m)}{2\left[\left(\frac{R_r}{s} + R_m\right) + j(X_r + X_m)\right]} \qquad (14)$$

and the voltage V_{a1} across the positive-sequence rotor circuit of Fig. 11 is as follows

$$V_{alr}=\frac{I_a}{3}\frac{\left[\left(\frac{R_{r1}}{s}+jX_{r1}\right)(R_{m1}+jX_{m1})\right]}{\left[\left(\frac{R_{r1}}{s}+R_{m1}\right)+j(X_{r1}+X_{m1})\right]}$$

$$=\frac{I_a}{3}\frac{\left[\frac{3}{2}\left(\frac{R_r}{s}+jX_r\right)\frac{3}{2}(R_m+jX_m)\right]}{\left[\frac{3}{2}\left(\frac{R_r}{s}+R_m\right)+\frac{3}{2}j(X_r+X_m)\right]} \quad (15)$$

It is observed that $V_f=V_{alr}$. By the same method, the negative-sequence rotor voltage of Fig. 11 is found identical to the backward field voltage of Fig. 16.

Lyons[2] has shown that the magnetizing reactance for the forward revolving zero-sequence field when saturation is neglected (with the system of symbols in this paper) is

$$\frac{X_{m0}}{2}=\frac{1}{9}\frac{(K_{w0})^2}{(K_{w1})^2}X_{m1} \quad (16)$$

By a similar method, the relationship between X_m and X_{m3}, Fig. 16, is

$$\frac{X_{m3}}{2}=\frac{1}{9}\frac{(K_{w3})^2}{(K_w)^2}\frac{X_m}{2} \quad (17)$$

Since $K_{w0}=K_{w3}$ and $K_{w1}=K_w$, equations 16 and 17 are combined to give:

$$\frac{X_{m0}}{X_{m3}}=\frac{2X_{m1}}{X_m} \quad (18)$$

By substituting $\frac{3}{2}X_m$ for X_{m1}, it is found that

$$\frac{X_{m0}}{2}=3\frac{X_{m3}}{2} \quad (19)$$

Leakage impedance of the bars of a squirrel cage rotor bar assigned to the forward revolving third space harmonic is

$$X_{r3}=X_r\left(\frac{CK_{w3}}{CK_w}\right)^2 \quad (20)$$

End ring leakage, skin effect, and the effect of skew are not considered here since these will not affect the relationship between X_{r0} and X_{r3}. Lyons[2] has shown that rotor bar leakage with skin effect neglected is

$$\frac{X_{r0}}{2}=\left(\frac{K_{w0}}{K_{w3}}\right)^2X_{r3} \quad (21)$$

Combining equations 20 and 21 gives:

$$\frac{X_{r0}}{2X_{r3}}=\frac{X_{r1}}{X_r} \quad (22)$$

Substituting $\frac{3}{2}X_r$ for X_{r1} gives:

$$\frac{X_{r0}}{2}=\frac{3X_{r3}}{2} \quad (23)$$

$R_{m0}/2$ and $R_{m3}/2$ are resistances included in their respective circuits to simulate core loss caused by third harmonics. Since the third space harmonic and zero-sequence fields are identical, $R_{m0}/2$ and $R_{m3}/2$ must be in such a ratio that they give the same loss.

By examining Figs. 11 and 16 and equation 19 it is seen that the current through $R_{m0}/2$ is one third the current through $R_{m1}/2$. Since Fig. 11 is on a per-phase basis of the 3-phase motor, the ratio of the constants is found as

$$(I_{m3})^2\frac{R_{m3}}{2}=\frac{I_{m0}}{3}^2\frac{3R_{m0}}{2}$$

$$\frac{R_{m0}}{2}=\frac{3R_{r3}}{2} \quad (24)$$

The voltage drop across the network representing the forward revolving zero-sequence field of Fig. 11 is the same as the voltage drop across the network representing the forward field for the third harmonic of Fig. 16.

Since the ratio of the currents is 1.3 and the ratio of the constants is 3 to 1.

$$\frac{I_{m3}}{3}\left(\frac{R_{m0}}{2}+\frac{jX_{m3}}{2}\right)=I_{m3}\left(\frac{R_{m3}}{2}+\frac{jX_{m3}}{2}\right)$$

Torques for the zero-sequence and third space harmonic field must be identical and the currents are in the ratio of 1 to 3. Therefore:

$$(I_{r0f})^2m\frac{R_{r0}}{2(3s-2)}=(3I_{r0f})^2\frac{R_{r3}}{2(3s-2)}$$

and

$$\frac{R_{r0}}{2}=\frac{3R_{r3}}{2} \quad (25)$$

The ratio has been established between the zero-sequence constants for Fig. 11 and the third space harmonic constants for Fig. 16.

$$\frac{X_{m0}}{2}=\frac{3X_{m3}}{2}\qquad\frac{R_{m0}}{2}=\frac{3R_{m3}}{2}$$

$$\frac{X_{r0}}{2}=\frac{3X_{r3}}{2}\qquad\frac{R_{r0}}{2}=\frac{3R_{r3}}{2} \quad (26)$$

It is now apparent that the stator impedance drop of Fig. 16 must equal the sum of the stator impedance drops of Fig. 11

$$I_a(R_s+jX_s)=\frac{I_a}{3}[(R_{s1}+R_{s2}+R_{s0})+$$
$$j(X_{s1}+X_{s2}+X_{s0})]$$

For this relation to be true, the following relations must hold:

$$R_s=\frac{(R_{s1}+R_{s2}+R_{s0})}{3} \quad (27)$$

$$X_s=\frac{(X_{s1}+X_{s2}+X_{s0})}{3} \quad (28)$$

When eddy-current and stray-load losses are neglected,

$R_{s1}=R_{s2}=R_{s0}$=ohmic resistance per phase

References

1. Symmetrical Components (book), C. F. Wagner, R. D. Evans. McGraw-Hill Book Company, Inc., New York, N. Y., 1933.

2. Standard Handbook for Electrical Engineers (book), A. E. Knowlton, editor. McGraw-Hill Book Company, Inc., eighth edition, 1949.

3. Application of the Method of Symmetrical Components (book), W. V. Lyons. McGraw-Hill Book Company, Inc., 1937.

4. Operation of 3-Phase Induction Motors on Unbalanced Voltages, J. E. Williams. *AIEE Transactions*, vol. 73, pt. III-A, Apr. 1954, pp. 125–33.

5. Some Aspects of the Application of Induction Motors to Aircraft, H. J. Braue. *Ibid.*, vol. 63, Oct. 1944, pp. 769–72; *discussion by* C. G. Veinott, p. 1476.

Discussion

G. R. Slemon (University of Toronto, Toronto, Ont., Canada): I would like to compliment Mr. Mitchell on his skillful use of equivalent circuit methods in this paper. There is a great educational advantage in considering the torque-producing effects separately, deriving an equivalent circuit for each, and then interconnecting these circuits to represent the particular form of unbalance. The interconnections of sequence circuits for series and shunt unbalances have been used extensively in the analysis of power systems but have been somewhat neglected in the literature on motor operation.

Since the use of symmetrical components assumes linearity, I would like to ask the author how he proposes to include the effects of nonlinearity, particularly of the magnetizing reactances, in the practical use of his circuits.

Interconnected equivalent circuits become somewhat cumbersome to calculate by hand methods. What practical methods would the author recommend for solution of these circuits?

J. T. Mitchell: Nonlinearity of the circuit values greatly complicates the evaluation of performance. It is very difficult to make accurate calculations for the saturated value of positive-, negative-, and zero-sequence magnetizing reactances even when the positive-, negative-, and zero-sequence voltages are known. Since a different magnetic path is encountered by each component mmf, approximations must be made. The magnetizing reactances change with load to complicate the problem further. Rotor bar skin effect can be approximated with a fair degree of accuracy.

Solution of the interconnected equivalent circuits by hand calculations is very laborious and the chance for errors is quite large. Therefore, calculations of complete performance curves is impractical even if the circuit values are assumed to be linear. Calculations for four load points will usually suffice for most applications. For hand calculations, the author recommends that unsaturated values for magnetizing reactances be used for locked rotor, minimum pull-up torque, and maximum torque. For the full load point, it will usually be necessary to make two trys. An estimate of the terminal voltages can be used to determine the sequence voltages. For the condition of one line open, the voltage on the open phase can be assumed

to be approximately 83% of the normal phase voltage. Saturated values for the magnetizing reactances can then be calculated (approximated) for this assumed condition. Estimates of full-load slip and skin effect are made and the equivalent circuit solved for more accurate values of the sequence voltages. The reactances are re-evaluated and used with a corrected slip value to solve the circuit for full-load performance.

The most practical way to obtain data for complete performance curves is by the use of a digital computer. If an analog computer is used, it must have means for providing negative resistance in the zero-sequence rotor circuit.

It must be emphasized that it is necessary to know the negative- and zero-sequence impedances of the system as viewed from the motor terminals. These impedances can have a serious effect on voltage and current unbalance and motor performance. Likewise, a power source of the same impedance must be used for test purposes if a reasonable agreement between test and calculated data is to be expected.

Graphical Analysis and Synthesis of Feedback Control Systems

I—Theory and Analysis

DUŠAN MITROVIĆ
NONMEMBER AIEE

IN THE DESIGN of high-performance feedback control systems it is necessary to consider simultaneously both the frequency and the transient response. Several methods have been developed to this end.[1,2] The method proposed in this paper also operates in terms of both frequency and time domains, but it still permits the entire process to be effected in the real domain, thus greatly simplifying the practical application.

Important advantages of the proposed method may be understood by the following reasoning: By use of Laplace transformation the problems of analysis and synthesis of linear feedback control systems are conveyed from the realm of differential equations to the algebraic domain. Owing to the difficulties inherent in the handling of algebraic equations, most of the existing methods are leaving even this domain, though it is not certain that the solution of engineering problems is thereby facilitated. The proposed method, on the contrary, approaches the problem algebraically, thus handling it directly in the form as obtained after applying the Laplace transformation. The most important advantages thereby accomplished are

All operations are effected in the real domain.

Coefficients of the characteristic equation can be adjusted without any calculation.

All roots of the characteristic equation can be estimated almost simultaneously.

This last property requires further explanation. The response of a system is governed by the complete set of all poles and zeros. Since the zeros are usually known, the only problem is in the poles. To get an idea of the response, however, it is necessary to know all the poles together. No method, therefore, which provides only for the solution of single roots, instead of all the roots of the characteristic equation simultaneously, would be suitable for the synthesis of feedback control systems.

An essential role in the proposed method is played by three curves. The problem is solved by studying the position of a working point in relation to these curves. It is not always necessary to plot all three of them. For instance, only one curve is sufficient for the solution of the stability problem. In the synthesis of feedback control systems only one or two curves are frequently sufficient. Their construction is quite easy. Moreover, in some cases one may use the curves constructed once, for all.

Application is not restricted only to the sphere covered by the examples. The method in general is suitable for application whenever it is required to examine how poles and zeros are affected by a change in parameters. Such, for instance, is the case of designing systems with multiple inputs. As regards nonlinear systems, the method is applicable to those cases in which one can validly assume that a transfer function based on the fundamental component of a nonsinusoidal waveform yields a sufficiently accurate result. The method can also be applied in examining active electrical networks, while some of its paragraphs may be of use to other methods.

Thus, for both linear and nonlinear systems the value of the method seems to lie in its simple and direct use of algebraic and graphical operations in the real domain and in its capabilities being almost equal to all other techniques together. The method enables the application of inverse Laplace transformation for obtaining the exact response.

Analysis and synthesis of sampled-data feedback control systems is also possible. Analysis of stability, the design procedure when the settling time or the relative damping ratio are specified, and system compensation are the most important applications in this field.

This being an entirely new method representing a whole in itself, the presentation has to be rather extensive. It is also necessary to include many examples. The detailed treatment of numerical examples may appear to be superfluous, but in the first presentation, at least, this is indispensable. In this discussion, if one discards all that serves the exclusive purpose of explanation, it will be readily seen that the practical application of the proposed method presents no difficulties.

Mathematical Bases

Suppose the closed-loop transfer function of a feedback system is given in the form

$$\frac{C}{E}(S)=\frac{p(s)}{f(s)} \qquad (1)$$

where $p(s)$ and $f(s)$ are polynomials. The behavior of a feedback system is determined by the poles and zeros of the closed-loop transfer function, i.e., by the zeros of the polynomials $p(s)$ and $f(s)$. It is, therefore, of particular interest to consider the roots of the equation

$$f(s)=a_n s^n+a_{n-1}s^{n-1}+\ldots+a_2 s^2+a_1 s+a_0=0 \qquad (2)$$

Paper 58-860, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted December 4, 1956; made available for printing April 22, 1958.

DUŠAN MITROVIĆ is with the University of Belgrade and the Boris Kidritch Institute, Belgrade, Yugoslavia.

Table I. Functions $\phi_k(\zeta)$

ζ	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
0.00	−1	0.0	+1.00	0.000	−1.0000	0.00000	+1.000000	0.0000000	−1.00000000	0.000000000	
0.05	−1	0.1	+0.99	−0.199	−0.9701	+0.29601	+0.940499	−0.3900599	−0.90149301	+ 0.430209301	
0.10	−1	0.2	+0.96	−0.392	−0.8816	+0.56832	+0.767936	−0.7219074	−0.62355456	+ 0.846518112	
0.15	−1	0.3	+0.91	−0.573	−0.7381	+0.79443	+0.492771	−0.9443813	−0.21646261	+ 1.000300083	
0.20	−1	0.4	+0.84	−0.736	−0.5456	+0.95424	+0.163904	−1.0198016	+0.24401654	+ 0.922194944	
0.25	−1	0.5	+0.75	−0.875	−0.3125	+1.03125	−0.203125	−0.9296875	+0.66796875	+ 0.595703125	
0.30	−1	0.6	+0.64	−0.984	−0.0496	+1.01376	−0.558656	−0.6785664	+0.96579584	+ 0.099088896	
0.35	−1	0.7	+0.51	−1.057	+0.2299	+0.89607	−0.857149	−0.2960657	+1.05439499	− 0.449010793	
0.40	−1	0.8	+0.36	−1.088	+0.5104	+0.67968	−1.054144	+0.1636352	+0.92323584	− 0.902223872	
0.45	−1	0.9	+0.19	−1.071	+0.7739	+0.37449	−1.110941	+0.6253569	+0.64811979	− 1.118664711	
0.50	−1	1.0	0.00	−1.000	+1.0000	0.00000	−1.000000	+1.0000000	0.00000000	− 1.000000000	
0.55	−1	1.1	−0.21	−0.869	+1.1659	−0.41349	−0.711061	+1.1956571	−0.60416181	− 0.531079109	
0.60	−1	1.2	−0.44	−0.673	+1.2464	−0.82368	−0.257984	+1.1333608	−1.10192896	+ 0.189053952	
0.65	−1	1.3	−0.69	−0.403	+1.2139	−1.17507	+0.313691	+0.7672717	−1.31114421	+ 0.937215773	
0.70	−1	1.4	−0.96	−0.056	+1.0384	−1.39776	+0.918454	+0.1119104	−1.07513856	+ 1.393283584	
0.75	−1	1.5	−1.25	+0.375	+0.6875	−1.40625	+1.421875	−0.7265625	−0.33203125	+ 1.324609375	
0.80	−1	1.6	−1.56	+0.896	+0.1264	−1.09824	+1.630784	−1.5110144	+0.78683904	+ 0.353071936	
0.85	−1	1.7	−1.89	+1.513	−0.6321	−0.35343	+1.282931	−1.8275527	+1.82990859	− 1.373091903	
0.90	−1	1.8	−2.24	+2.232	−1.7776	+0.96768	+0.035776	−1.0320768	+1.82196224	− 2.247455232	
0.95	−1	1.9	−2.61	+3.059	−3.2021	+3.02499	−2.545381	+1.8112339	−0.89596341	− 0.108903421	
1.00	−1	2.0	−3.00	+4.000	−5.0000	+5.00000	−7.000000	+8.0000000	−9.00000000	+10.000000000	

which is known as the characteristic equation.

Let be given equation 2 where all coefficients are real and $a_n > 0$. After substituting variables ξ and η for a_1 and a_0, equation 2 becomes

$$a_n s^n + a_{n-1} s^{n-1} + \ldots + a_2 s^2 + \xi s + \eta = \psi(s) + \xi s + \eta = 0 \quad (3)$$

Now, transform equation 3 by introducing

$$s = \omega_n e^{j(\pi/2+\theta)} = -\omega_n \sin\theta + j\omega_n \cos\theta = -\omega_n \zeta + j\omega_n \sqrt{1-\zeta^2} \quad (4)$$

where θ and ζ are confined to the intervals $0 \leq \theta \leq \pi/2, 0 \leq \zeta \leq 1$.

After the real and the imaginary part are each set equal to zero, a locus of points in the $0\xi\eta$ plane corresponding to the roots with ζ constant is obtained. Its equations are

$$\xi = a_2\phi_2(\zeta)\omega_n + a_3\phi_3(\zeta)\omega_n^2 + \ldots + a_n\phi_n(\zeta)\omega_n^{n-1}$$

$$\eta = -\omega_n^2[a_2\phi_1(\zeta) + a_3\phi_2(\zeta)\omega_n + \ldots + a_n\phi_{n-1}(\zeta)\omega_n^{n-2}] \quad (5)$$

The variable ζ denotes the damping coefficient, while ω_n represents the undamped natural frequency. The functions $\phi_k(\zeta)$ are polynomials in ζ which may be obtained by successive application of the relation

$$\phi_k(\zeta) = -[2\zeta\phi_{k-1}(\zeta) + \phi_{k-2}(\zeta)] \quad (6)$$

with $\phi_0(\zeta) = 0$ and $\phi_1(\zeta) = -1$. As an example, the first eight functions $\phi_k(\zeta)$ will be determined. From equation 6 the following are easily obtained:

$$\phi_1(\zeta) = -1 \qquad \phi_5(\zeta) = -16\zeta^4 + 12\zeta^2 - 1$$
$$\phi_2(\zeta) = 2\zeta \qquad \phi_6(\zeta) = 32\zeta^5 - 32\zeta^3 + 6\zeta$$
$$\phi_3(\zeta) = -4\zeta^2 + 1 \qquad \phi_7(\zeta) = -64\zeta^6 + 80\zeta^4 - 24\zeta^2 + 1$$
$$\phi_4(\zeta) = 8\zeta^3 - 4\zeta \qquad \phi_8(\zeta) = 128\zeta^7 - 192\zeta^5 + 80\zeta^3 - 8\zeta \quad (7)$$

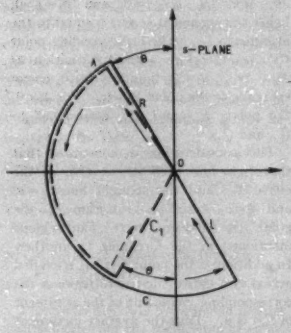


Fig. 1. Contours c and c_1 enclosing all roots with damping coefficients greater than $\zeta = \sin\theta$

Successive steps 4, 5, and 6 are more completely explained by an example in the Appendix.

Functions $\phi_k(\zeta)$ are affected neither by the coefficients nor the degree of equation 2. They may therefore be computed once for all and collected in a table which may easily be supplemented or extended; see Table I.

With the use of such a table, equations 5 of the locus corresponding to the specified damping coefficient ζ can be written practically without any calculation. The right-hand members of equations 5 are then obtained as polynomials in ω_n and the construction of the corresponding curve Γ_ζ involves no difficulties.

As will shortly be seen, the shape of the locus corresponding to the roots with the prescribed ζ indicates whether the damping coefficients of all roots of equation 2 are greater than the specified ζ. In the special case of $\zeta = 0$ the corresponding curve indicates whether the system is stable, while the curve Γ_1 obtained for $\zeta = 1$ serves to determine whether all roots are real and negative, i.e., whether the system is overdamped. To prove this, the graphical conditions for the damping coefficients of all roots of equation 2 will be derived to be greater than the specified ζ.

To this end, consider, in Fig. 1, two closed contours C and C_1 with infinite radius R. For the damping coefficients of all roots of the algebraic equation 2 to be greater than the specified ζ all roots must obviously be enclosed by the contour C_1. Contours C and C_1, however, are equivalent in this respect. This implies that, if all roots of equation 2 are enclosed by the contour C they will also be enclosed by the contour C_1, and conversely. Assuming $f(s)$ on these contours to be different from zero, first consider the contour C. Since the polynomial $f(s)$ is a holomorphe function on and within the contour C, it follows from Cauchy's theorem that all roots of equation 2 will be enclosed by the contour C if the increment of the argument of $f(s)$ on the closed contour C is equal to $2n\pi$. Since for s very large, $f(s)$ behaves as $a_n s^n$, the increment of the argument of $f(s)$ on the semicircular part of the contour C will be $n\pi$ and the preceding condition may also be expressed in the following terms: All roots of equation 2 will be enclosed by the contour C if the increment in the argument of $f(s)$ on the straight line l from $R\zeta - jR\sqrt{1-\zeta^2}$ to $-R\zeta + jR\sqrt{1-\zeta^2}$ is $n\pi$ exactly. To check on this condition being satisfied, one should first substitute $s = -\omega_n\zeta + j\omega_n\sqrt{1-\zeta^2}$ in $f(s)$. This substitution, however, can be effected more easily in an indirect way. From equations 2 and 3 it follows that

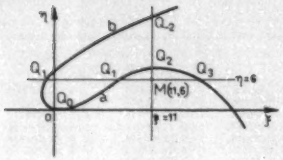


Fig. 2. Curve $\Gamma_{0.4}$ and the working point M for equation

$$s^3+6s^2+11s+6=0$$

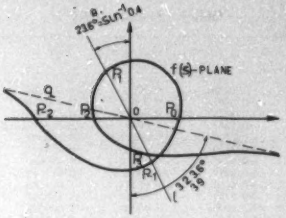


Fig. 3. Approximate transform of the straight line l in the f(s) plane for the example of Fig. 2

$$f(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2}) = \psi(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})+ a_1(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})+a_0 \quad (8)$$

Equation 3 is satisfied for every point $P(\xi,\eta)$ of the curve 5. Therefore

$$0=\psi(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})+ \xi(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})+\eta \quad (9)$$

Substraction of equation 9 from equation 8 finally yields the results of substitution in the form

$$f(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2}) = a_0-\eta+(a_1-\xi)(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2}) \quad (10)$$

where η and ξ can easily be determined with the help of equation 5.

Relation 10 may be of use for a variety of purposes. It will be applied in the example that follows to determine the increment of the argument of $f(s)$ on the straight line l. Consider

$$f(s)=s^3+6s^2+11s+6=0 \quad (11)$$

According to equations 5 and 6 or equation 7, the equations of the locus corresponding to the constant damping coefficient $\zeta=0.4$ are

$$\xi=4.8\omega_n+0.36\omega_n^2$$
$$\eta=6\omega_n^2-0.8\omega_n^3 \quad (12)$$

These equations could as well have been written directly using Table I.

Let the curve $\Gamma_{0.4}$ defined by equations 12 as well as the straight lines $\eta=a_0=6$ and $\xi=a_1=11$ be constructed. This is shown in Fig. 2 (not drawn to scale). In Fig. 3 the plane $f(s)$ is divided into four regions separated by the real axis and the straight line l corresponding to $\zeta=0.4$.

First, examine the case $\omega_n=0$. Note that $s=0$, $\xi=0$, $\eta=0$ and $f(s)=f(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})=a_0=6$. Thus the argument of $f(s)$ for $\omega_n=0$ is equal to zero. The point $f(s)$ has a similar position as P_0 in Fig. 3. The corresponding point in Fig. 2 is denoted by Q_0.

If ω_n is allowed to increase, the curve $\Gamma_{0.4}$ passes through points Q_1, Q_2, and Q_3. At the same time, the point $f(s)$ crosses the line l or the real axis in points P_1, P_2, and P_3, as evident from equation 10. For Q_1, $a_0-\eta=0$ and $a_1-\xi>0$; hence the argument of $f(s)$ is equal to the argument of s. The corresponding point $f(s)$ has therefore a similar position as P_1. At Q_2, $a_1-\xi=0$ and $a_0-\eta<0$, corresponding to the point P_2, etc. For $\omega_n<0$, the points P_{-1} and P_{-2} correspond to Q_{-1} and Q_{-2}.

This consideration demonstrates that the origin and each intersection of the curve $\Gamma_{0.4}$ and the straight lines $\eta=a_0$ and $\xi=a_1$ entail a transition of the point $f(s)$ from one of the four regions mentioned in the foregoing to another. In addition, if the point $f(s)$ has travelled across two entire successive regions, the corresponding increment of the argument of $f(s)$ is π. Thus the average increment corresponding to an intersection is $\pi/2$. A total increment of $n\pi$ would require $2n-1$ points of intersection, or $2n-2$ in special cases (since the origin plays the same role as an intersection).

Fig. 2 may also be used to inspect the sign of the increment of the argument of $f(s)$. In example 11, the increment is always positive for ω_n between $-\infty$ and $+\infty$. Generally, Fig. 2 is useful in constructing the approximate shape of the transform of the line l in the $f(s)$ plane. For this purpose it should be noted that for ω_n very large $f(s)$ behaves as $a_n s^n$. Therefore the asymptotal direction of the transform is $n(\pi/2+\theta)$. In Fig. 3 this direction is represented by the dashed line q.

As a further example of the approximate plotting of the transform of the line l in the $f(s)$ plane, consider the following algebraic equation:

$$s^3+6s^2+60s+4=0 \quad (13)$$

Equations of the locus for $\zeta=0.4$ are the same as in the preceding example except that the point $M(60.4)$ has a different location; this is illustrated in Fig. 4. The approximate transform of the line l in the $f(s)$ plane is plotted in Fig. 5, which obviously can be constructed by simply inspecting Fig. 4. In this example the total increment of the argument of $f(s)$ is π. Only one root, therefore, has a damping coefficient greater than 0.4.

Now, consider equations 12. To find the points of intersection of this curve $\Gamma_{0.4}$ with the straight lines $\eta=a_0$ and $\xi=a_1$, first, $\eta=6$ and $\xi=11$ must be substituted in equation 12, and then each of the resulting equations must be solved for ω_n. Evidently, the largest possible number of points of intersection is $2+3=5=2n-1$. The same number of points of intersection is required for the total increment of the argument of $f(s)$ to be $n\pi$. It may be concluded, therefore, that no point of intersection is allowed to result in a negative increment of the argument of $f(s)$. This is true for every n.

To avoid the construction of the curve Γ_ζ for $\omega_n<0$, consider the contour C_1 (Fig. 1). All roots of equation 2 will be inside C_1 if the increment of the argument of $f(s)$ on the closed contour C_1 is exactly $2n\pi$. On the circular arc of infinite radius, however, this increment is $2n(\pi/2-\theta)$. Hence the increment on two half-lines must be $2n\pi-2n(\pi/2-\theta)=2n(\pi/2+\theta)$. On account of symmetry this condition can be expressed in a simpler manner: All roots of equation 2 will be inside the contour C_1 if the increment of the argument of $f(s)$ on the half-line $0A$ from 0 to $-R\zeta+jR\sqrt{1-\zeta^2}$ is $n(\pi/2+\theta)$ exactly. To check on this condition being satisfied one may use the same procedure as in the preceding example except that the part of the curve Γ_ζ for $\omega_n<0$ need no more be taken into consideration. From the branch a in Fig. 2, for instance, the part $P_0P_1P_2P_3g$ of the curve in Fig. 3 corresponding to $\omega_n>0$ is obtained. With the help of this part it is concluded that in the case under consideration the increment of the argument of $f(s)$ on the half-line $0A$ is $3\pi/2+3\theta=n(\pi/2+\theta)$. All roots are therefore inside the contour C_1.

By inspecting Figs. 2 and 3, while bearing in mind that no point of intersection is allowed to result in a negative increment of the argument of $f(s)$ one concludes that the straight lines $\eta=a_0$ and $\xi=a_1$ must be cut alternately and that the line $\eta=a_0$ must be cut first. On the other hand, a positive initial increment between P_0 and P_1 requires at the first point of intersection $a_0-\eta=0$ and $a_1-\xi>0$, which is possible only if the point $M(a_1, a_0)$ lies in the first quadrant.

It follows further from the foregoing discussion that there is a relationship

between the number of points of intersection and the increment of the argument of $f(s)$. Therefore, determine the number of points of intersection m which, if other conditions are met, will yield an increment in the argument of $f(s)$, amounting to $n(\pi/2+\theta)$.

The following correlation between the points of intersection and the increments of the argument of $f(s)$ may be deduced from Figs. 2 and 3:

Point	Increment
Q_0	$\pi/2+\theta$
Q_1	$\pi/2-\theta$
Q_2	$\pi/2+\theta$
⋮	
Q_m	$\pi/2-(-1)^m\theta$

Between points Q_0 and Q_m the corresponding increment is $m\pi/2+\theta$ for an odd m and $m\pi/2$ for an even m. The expression $m\pi/2+[1-(-1)^m]\theta/2$ obviously comprises both cases. Fig. 3, on the other hand, shows that the increment $n(\pi/2+\theta)$ is always greater than the increment between Q_0 and Q_m. The following inequality therefore holds

$$m\frac{\pi}{2}+\frac{1-(-1)^m}{2}\theta<n\left(\frac{\pi}{2}+\theta\right) \qquad (14)$$

Evidently, the increment $n(\pi/2+\theta)$ can be realized only if m is the largest integer satisfying inequality 14.

The general conclusion resulting from the preceding considerations may be expressed in terms of the following necessary and sufficient conditions[3]: The damping coefficient of every root of the characteristic equation 2 (more precisely, the damping coefficient of every oscillatory component in the step function response), will be greater than the specified ζ if the point $M(a_1, a_0)$ lies in the first quadrant and if, while ω_n varies from 0 to $+\infty$ the curve Γ_ζ defined by equations 5 alternately cuts lines $\eta=a_0$ and $\xi=a_1$ with the provision that the line $\eta=a_0$ is cut first and that the total number of points of intersection is m where m is the largest integer satisfying inequality 14.

Construction of Curves Γ_ζ for $\zeta=0$ and ζ Between 0 and 1

The application of the curves Γ_ζ is much wider than may be assumed from the previously derived criterion. They represent the main tools for the analysis and synthesis of feedback control systems by the proposed method. There is, however, an essential difference in their application, depending on the value of ζ, for which they are constructed. It will therefore be necessary in future consideration to distinguish between the following three types of curves:

1. Γ_0 = curve constructed for $\zeta = 0$.
2. Γ_ζ = curve for ζ between 0 and 1.
3. Γ_1 = curve for $\zeta = 1$.

It is usually not necessary to use all the three curves, depending on the particular problem and the stage to which its solution has progressed. It would therefore be difficult to say which of the three curves is the most important one for application. There is, however, no difference as regards the construction of the curves Γ_0 and Γ_ζ and a special instruction will be necessary only for the curve Γ_1. This will be given in the next section.

To construct the curve Γ_ζ, one first writes equations 5 using the functions $\phi_k(\zeta)$, given in Table I, whereby ξ and η assume the forms of polynomials in ω_n. Assigning, then, different values to ω_n, one obtains the co-ordinates of the points on the curve corresponding to the selected ζ. This procedure is easy and obvious, but in order to simplify it still further it should be noted that equations 5 may be written in the form

$$\xi=a_2\psi_2(\zeta,\omega_n)+a_3\psi_3(\zeta,\omega_n)+\ldots+a_n\psi_n(\zeta,\omega_n)$$
$$\eta=-\omega_n\zeta[a_2\psi_1(\zeta,\omega_n)+a_3\psi_2(\zeta,\omega_n)+\ldots+a_n\psi_{n-1}(\zeta,\omega_n)] \qquad (15)$$

where $\psi_k(\zeta,\omega_n)$ denotes the function

$$\psi_k(\zeta,\omega_n)=\phi_k(\zeta)\omega_n^{k-1} \qquad (16)$$

Functions $\psi_k(\zeta,\omega_n)$ are affected neither by the degree nor the coefficients of equation 2. They can therefore be computed once for all, and collected in a common table in which the values of the functions $\psi_1, \psi_2, \psi_3 \ldots$ for one value of ζ and ω_n would appear in the same row. Computation of points on the curve would thus be reduced to a cumulative multiplication of the tabulated values and the coefficients a_k. If a desk machine is available it would be necessary to record only the final results.

The table of functions need not necessarily be as extensive as it might seem at first glance. If the curve to be plotted is intended for the analysis of a feedback system, the values of the damping coefficient need not be given with greater resolution than 0.05. Moreover, in cases $a_n=1$, substitution $s=a_{n-1}p$ or a similar one results in the values of ω_n between 0 and 1 for all roots.

Since the foregoing convenient substitution is in very frequent use, it should be discussed more thoroughly. Curves Γ_0, Γ_1 and Γ_ζ, corresponding to the characteristic equation 2, are often inconvenient as regards the scale. The determination of large-magnitude real roots, as will be shown later, is not sufficiently accurate, which may cause difficulties in solving other problems. Under such circumstances a transformation of equation 2 by setting $s=cp$ becomes self-evident. In cases of $a_n=1$ it is often most convenient if c is of the same order as or equal to the coefficient a_{n-1}. By this substitution the absolute values of all real roots are made to be less than unity. In this way the precision of results is increased and, as has been mentioned previously, the use of the table of functions $\psi_k(\zeta,\omega_n)$ is made possible.

It is important to note that with the help of the points of the curves Γ_0, Γ_1, and Γ_ζ one can obtain with practically no computation the points on the curves Γ_{0t}, Γ_{1t}, and $\Gamma_{\zeta t}$, corresponding to the transformed characteristic equation, and conversely. This may be demonstrated by substituting in equation 2

$$s=cp=c(-\omega_{nt}\zeta+j\omega_{nt}\sqrt{1-\zeta^2}) \qquad (17)$$

The transformed equation

$$a_n p^n+\frac{a_{n-1}}{c}p^{n-1}+\ldots+\frac{a_2}{c^{n-2}}p^2+\frac{a_1}{c^{n-1}}p+\frac{a_0}{c^n}=0 \qquad (18)$$

is thus obtained. The corresponding curves Γ_{0t}, Γ_{1t}, and $\Gamma_{\zeta t}$ are determined by the equations

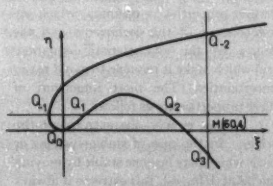


Fig. 4. Curve Γ_{0-4} and the working point M for equation

$$s^3+6s^2+60s+4=0$$

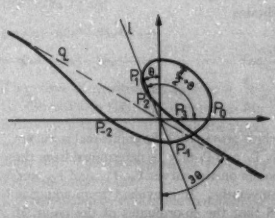


Fig. 5. Approximate transform of the straight line l in the f(s) plane for the example of Fig. 4

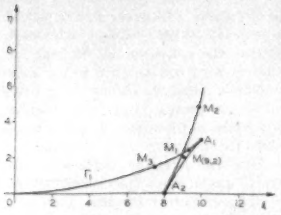


Fig. 6. Curve Γ_1 and the working point M for equation

$$s^4+6s^3+12s^2+(16+K)s+Kb=0$$

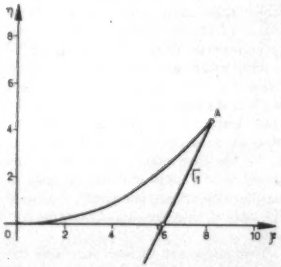


Fig. 7. Curve Γ_1 for equation

$$s^3+5s^2+6s+K=0$$

$$\xi_t=\frac{a_2}{c^{n-2}}\phi_2(\zeta)\omega_n t+\frac{a_3}{c^{n-3}}\phi_3(\zeta)\omega_n t^2+\ldots+a_n\phi_n(\zeta)\omega_n t^{n-1}$$

$$\eta_t=-\omega_n t^2\left[\frac{a_2}{c^{n-2}}\phi_1(\zeta)+\frac{a_3}{c^{n-3}}\phi_2(\zeta)\omega_n t+\ldots+a_n\phi_{n-1}(\zeta)\omega_n t^{n-2}\right] \quad (19)$$

From equation 17 it follows, however, that

$$\omega_n t=\frac{\omega_n}{c} \quad (20)$$

Substituting equation 20 in equation 19, and taking into account equation 15, one obtains

$$\xi_t=\frac{\xi}{c^{n-1}}$$

$$\eta_t=\frac{\eta}{c^n} \quad (21)$$

which proves that the points on the curves Γ_0, Γ_t, and Γ_ζ may be determined from the points on the curves Γ_{0t}, Γ_{1t}, and $\Gamma_{\zeta t}$, and conversely. It is evident from equation 18 that the co-ordinates of the point M should also be changed in accordance with equations 21.

The practical application of this transformation will be demonstrated in a later discussion. It is effective also in cases when the table of functions $\psi(\zeta, \omega_n)$ for ω_n between 0 and 1 is used.

Returning to the general discussion on construction of the curves Γ_ζ it should be noted that another particular value of ζ deserves special attention, namely, for $\zeta=0.5$ the functions $\phi_k(\zeta)$ assume the values -1, 1, or 0, as evident from equation 6 or equation 7. Therefore the equations of the locus of points corresponding to the roots whose damping coefficient is $\zeta=0.5$ take the very simple form

$$\xi=a_2\omega_n-a_4\omega_n^3+a_6\omega_n^4-a_7\omega_n^6+\ldots$$

$$\eta=a_2\omega_n^2-a_3\omega_n^3+a_5\omega_n^4-a_6\omega_n^5+\ldots \quad (22)$$

Since this curve $\Gamma_{0.5}$ divides the underdamped region into two parts, corresponding to $\zeta<0.5$ and $\zeta>0.5$ (if both exist), it finds very useful application in feedback control systems.

As regards the construction of the curves Γ_ζ, it should be pointed out that a cusp may appear only exceptionally and that the construction of these curves by the use of data computed in the form of Table I involves no difficulties. In case the table of functions $\psi_k(\zeta,\omega_n)$ is used, it is important to note that the calculation of points on the curve Γ_ζ increases only slightly in complexity as the degree of the characteristic equation is raised.

Construction of Locus Corresponding to Double-Negative Real Roots

To obtain the equations of the curve Γ_1, set in 5 $\zeta=1$. On account of 6 or 7 this yields

$$\xi=2a_2\sigma-3a_3\sigma^2+4a_4\sigma^3-5a_5\sigma^4+\ldots$$

$$\eta=a_2\sigma^2-2a_3\sigma^3+3a_4\sigma^4-4a_5\sigma^5+\ldots \quad (23)$$

Furthermore, for $\zeta=1$, equation 4 gives

$$s=-\sigma \quad (24)$$

which implies that equations 23 represent the locus of points corresponding to the double real roots of the characteristic equation 2.

Curves Γ_1 defined by equations 23 have certain properties in common, which are independent of the coefficients and the degree of the characteristic equation, and which make it possible to avoid many uncertainties. The most important of these properties are as follows:

1. There is only one extreme at the origin. In the case of stable systems or those which may become stable by moving the point $M(a_1, a_0)$, this extreme is always a minimum.

2. The curves Γ_1 never possess points of inflection.

3. The cusps may be found by determining the odd-order real roots of the algebraic equation

$$f^n(-\sigma)=\frac{d\xi}{d\sigma}=0 \quad (25)$$

and substituting the resulting values in 23. The maximum number of cusps is equal to the degree of equation 25, which is $n-2$. At first glance it may seem that we are again confronted with the difficulty of solving algebraic equations. For several reasons, however, this is not true. First, the aim is not to find the cusps but to determine how to connect the calculated points, wherefore it is sufficient to know the other properties of the curves Γ_1 and to bear in mind that the number of cusps may only be one of the following: $n-2$, $n-4$, $n-6$ .... The approximate plotting of the cusps therefore presents no difficulty. If, nevertheless, a more exact determination of these points should be desirable, the problem would still not be too difficult, since the degree of equation 25 is less by 2 than the degree of the characteristic equation. Moreover, as will later be demonstrated, the locus of points corresponding to the double real roots may be successfully used for the determination of real roots.

4. The curve Γ_1 has a total of $m+1$ branches, where m is the number of cusps. Hence, the maximum number of branches is $n-2+1=n-1$. A branch is the part of the curve Γ_1 whose curvature has always the same sign. In passing from one branch to the other, the curve crosses from one side of the tangent to the other. This entails a change in the sign of the derivative $d\xi/d\sigma$ whose magnitude in addition indicates the distance from the cusp.

5. From equation 23 one easily obtains the derivative

$$\eta'=\frac{d\eta}{d\sigma}=\sigma \quad (26)$$

which implies that the slope of the tangent at every point of the curve Γ_1 is equal to the parameter σ. Hence, the calculation of points gives not only the co-ordinates of the points but also the corresponding slopes, which greatly simplifies the con-

Table II. Calculation of Points of Curve Γ_1 Corresponding to $s^3+5s^2+6s+K=0$

ω_n	ξ	η
0.0	0.0	0.0
0.5	8.0	1.668
1.0	10.0	3.0
1.5	9.0	1.668
2.0	8.0	0.0
3.0	18.0	27.0

struction. Moreover, it is obvious from equation 26 that, starting from the origin, the slope of the curve constantly increases.

On the basis of the foregoing properties, one may summarize the following general instructions for the construction of curves:

1. If the characteristic equation is of the fourth degree or less, determine first the cusps by the use of equations 25 and 23 and then plot them together with their tangents. This, in most cases, will be sufficient for an approximate construction. If there are no cusps or greater accuracy is required, determine first the points of intersection of the curve and the co-ordinate axes by setting in equation 23 $\xi=0$ or $\eta=0$ and then calculate the other points by substituting the assumed values for σ in equation 23.

2. If the characteristic equation is of the fifth degree or higher, proceed immediately with the computation of points substituting in equation 23 the assumed values for σ. At the same time, one should bear in mind the five properties mentioned earlier, and draw tangents whenever any doubt should arise. Furthermore, in addition to ξ and η, it is advisable also to calculate $d\xi/d\sigma$ or at least to consider its sign. A change in the sign of this derivative means that the point has crossed to the next branch and that the curve Γ_1 has crossed to the other side of the tangent. Also, the less the value of this derivative, the nearer the calculated point to the cusp.

Take two examples: The first is to construct the curve Γ_1 corresponding to the characteristic equation

$$f(s)=s^4+6s^3+12s^2+(16+K)s+Kb=0 \quad (27)$$

where K and b are variable parameters. On account of equation 23 the equations of the curve Γ_1 are obtained:

$$\xi=24\sigma-18\sigma^2+4\sigma^3$$

$$\eta=12\sigma^2-12\sigma^3+3\sigma^4 \quad (28)$$

To find the cusps, the following equation is to be solved:

$$f'(-\sigma)=12\sigma^2-36\sigma+24=0$$

The solutions are $\sigma_1=1$ and $\sigma_2=2$; the corresponding cusps due to equation 28 are $A_1(10,3)$ and $A_2(8,0)$. Once these points are plotted in Fig. 6 and the corresponding tangents constructed, one can draw the approximate sketch of the curve Γ_1 bearing in mind properties 1, 2, 3, and 5. If greater accuracy is required, intermediate points are to be computed as in Table II.

Table III. Calculations of Points of Curve Γ_1 Corresponding to $s^5+6s^4+11s^3+6s^2+s+1=0$

ω_R	ξ	η	$d\xi/d\sigma$
0.1	0.89	0.04	6.1
0.2	1.26	0.091	1.5
0.3	1.24	0.082	−1.9
0.4	0.928	−0.028	−4.2
0.7	−0.738	−0.956	−5.8
1.0	−2.0	−2.0	−2.0
1.3	−1.72	−1.50	3.9
2.0	4.0	8.0	8.0
2.5	3.44	6.25	−15.5
3.0	−18.0	−54.0	−78.0

As a second example, construct the curve Γ_1 corresponding to the characteristic equation

$$s^5+5s^3+6s+K=0 \quad (29)$$

The curve Γ_1 is determined by equations

$$\xi=10\sigma-3\sigma^2$$

$$\eta=5\sigma^3-2\sigma^5 \quad (30)$$

and the cusp by

$$6\sigma-10=0 \quad (31)$$

The solution $\sigma=5/3$ substituted in equation 30 gives the point $A(25/3,\ 125/27)$ which should be plotted together with the tangent. If the intersection B is also determined by the use of equation $2\sigma^3-5\sigma^5=0$, and if its tangent is drawn, nothing more will be required in many cases. The shape of Γ_1 is shown in Fig. 7.

Determination of Roots

The locus of points corresponding to the double real negative roots, equation 23, may be used for graphical determination of negative real roots. (It can also be used to determine positive roots provided the part of the curve for $\sigma<0$ is also plotted.) This procedure is mainly based on relation 26 which is independent of the degree and the coefficients of the characteristic equation 2. The following statement holds:

The real roots of the algebraic equation 2 are equal to the opposite signed slopes of the tangents from $M(a_1, a_0)$ to the curve Γ_1.

To prove this, write the equation of the tangent to the curve Γ_1. Owing to equation 26 this equation is

$$\eta-\eta_1-\sigma_1(\xi-\xi_1)=0 \quad (32)$$

Here ξ and η are variable co-ordinates on the tangent, while ξ_1 and η_1 are the co-ordinates of the point on the curve Γ_1 determined by σ_1. If, however, the point $M(a_1, a_0)$ lies on tangent 32, the following relation must hold:

$$a_0-\eta_1-\sigma_1(a_1-\xi_1)=0 \quad (33)$$

On account of equation 23, 10 and $\zeta=1$, this implies that the equation

$$f(-\sigma_1)=0 \quad (34)$$

must also be satisfied.

Thus, if the point $M(a_1, a_0)$ lies on the tangent to the curve Γ_1, the value of its slope with opposite sign satisfies equation 2. The foregoing statement is thereby proved. It follows immediately that there are as many negative real roots of equation 2 as there are tangents to the curve Γ_1 passing through the point $M(a_1, a_0)$. If this curve is plotted also for $\omega_R<0$, the number of tangents then determines the total number of real roots. When dealing with linear feedback systems, however, the latter part of the curve Γ_1 is not necessary.

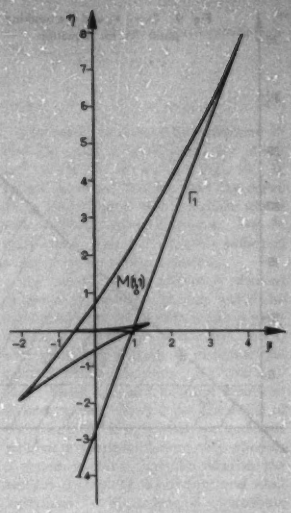


Fig. 8. Curve Γ_1 and the working point M for equation

$$s^5+6s^4+11s^3+6s^2+s+1=0$$

As an example, determine the negative real roots of the equation

$$s^5+6s^4+11s^3+6s^2+s+1=0 \quad (35)$$

On account of equation 23 the equations of the curve Γ_1 are

$$\xi=12\sigma-33\sigma^2+24\sigma^3-5\sigma^4$$

$$\eta=6\sigma^2-22\sigma^3+18\sigma^4-4\sigma^5 \quad (36)$$

The curve is constructed in Fig. 8 with the use of Table III. It follows from Fig. 8 that three tangents to Γ_1 with the slopes $\sigma_1=1$, $\sigma_2=2.5$, $\sigma_3=2.7$ pass through $M(1,\ 1)$. The corresponding roots are -1, -2.5 and -2.7.

It may often be impossible to determine accurately the large-magnitude roots by drawing tangents. This is mostly due to the lack of space for the corresponding part of the curve to be plotted. This

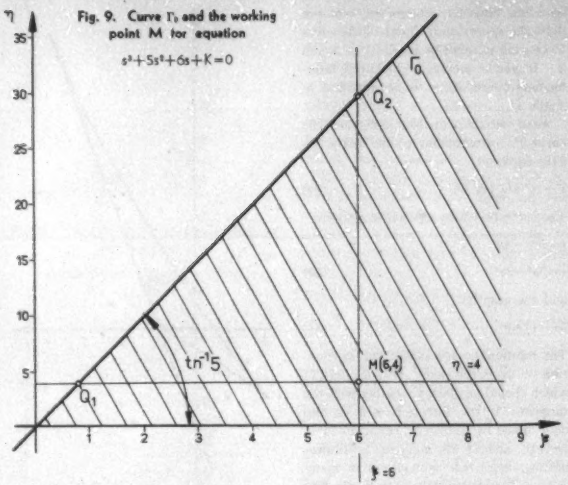


Fig. 9. Curve Γ_0 and the working point M for equation

$$s^5+5s^4+6s+K=0$$

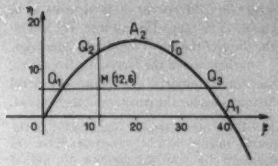


Fig. 10. Curve Γ_0 and the working point M for equation

$$s^4+5s^3+8s^2+12s+6=0$$

difficulty may be circumvented, however, by plotting the tangents themselves instead of the remote points. To show this, introduce the co-ordinates ξ_1 and η_1 from equation 23 in equation 32. The equation of the tangent to the curve Γ_1 then becomes

$$\eta-\sigma\xi+a_2\sigma^2-a_3\sigma^3+a_4\sigma^4-\ldots=0 \qquad (37)$$

A comparison of equations 37 and 2 reveals that the equation of the tangent to the curve Γ_1 may simply be written by setting $s=-\sigma$ in the characteristic equation, and taking coefficients a_1 and a_0 as variable co-ordinates ξ and η.

Tangents 37 cut the ξ axis at points

$$\xi=a_1\sigma-a_2\sigma^2+a_4\sigma^3-\ldots \qquad (38)$$

while their slope is σ. Hence, although the remote part of the curve is not constructed, it is possible to construct the tangents for several values of the parameter σ and to cover the region of interest. The corresponding root in that case is determined by interpolation between two consecutive tangents.

Starting from the fact that all negative real roots can easily be determined graphically, it is not difficult to demonstrate that one pair of conjugate complex roots can also be determined, under the condition that this pair is the only one and that all real roots of the given equation are negative. From the point of view of application to feedback control systems these conditions are not restrictive, since the calculation of roots is reasonable only if systems are stable, i.e., if all real roots of their characteristic equations are negative. Furthermore, a vast number of feedback control systems are, or can be, designed to be described by a closed-loop transfer function with only one pair of conjugate complex poles (control poles).

Under such circumstances the determination of roots will proceed in the following manner: All the $n-2$ real negative roots $-p_1, -p_2 \ldots -p_n$ are first determined by constructing the tangents. Since, according to equation 4, the absolute value of the complex root is ω_n and the real part is $-\omega_n\zeta$ the following equations hold:

$$p_1p_2\ldots p_n=s\omega_n^2=a_0 \qquad (39)$$

$$p_1+p_2+\ldots+p_{n-2}+2\omega_n\zeta=a_{n-1} \qquad (40)$$

With a simple calculation one obtains from these equations first ω_n and then ζ.

Equation 39 may be used also in cases when all roots are known except a real one. If only two real negative roots are unknown, they may also be determined with the help of equations 39 and 40.

In synthetising feedback control systems, one assumes a pair of control poles by selecting the point M on the curve Γ_ζ, which implies that one pair of conjugate complex roots is known. In that case the curve Γ_1 and equations similar to 39 and 40 permit another pair of conjugate complex roots of the characteristic equation as well as all real roots to be determined.

It should be stressed that with the aid of the table of functions $\psi_k(\zeta, \omega_n)$ defined by equation 16, and by applying the trial and error method, it is possible to determine all complex roots of the characteristic equation. The procedure is adapted for use with a desk machine capable of performing cumulative multiplications. The calculation process can be followed graphically in the same $0\xi\eta$ plane in which the curves $\Gamma\zeta$ are constructed and the determination of complex roots with the aid of the table of functions $\psi_k(\zeta,\omega_n)$ is therefore particularly suitable as an addition to the analysis and synthesis of feedback control systems by the method proposed in this paper. As it is impossible to give here the necessary tables, the determination of complex roots will be published elsewhere.

Stability

The condition of stability is frequently expressed in the following terms: A system will be stable if all roots of its characteristic equation possess negative real parts. Assuming $\omega_n>0$, however, this condition can obviously be formulated in another less common way: A system will be stable if the damping coefficient of every root of its characteristic equation is greater than zero. The problem of stability is thus only a special case of the conditions derived in the section "Mathematical Bases." Therefore set $\zeta=0$ in equation 5, which on account of equation 6 or equations 7 results in

$$\xi=a_2\omega^2-a_4\omega^4+a_6\omega^6-\ldots$$

$$\eta=a_3\omega^3-a_5\omega^5+a_7\omega^7-\ldots \qquad (41)$$

Substitution $\zeta=0$ in equation 4, on the other hand, yields

$$s=j\omega_n=j\omega \qquad (42)$$

Equations 41 thus represent the locus of points corresponding to purely imaginary roots of equation 2.

For $\zeta=0$, equation 4 gives $\theta=0$, which when substituted in inequality 14 results in $m=n-1$. Conditions of stability then read as follows:

A feedback control system described by the characteristic equation 2 will be

stable if the point $M(a_1, a_0)$ lies in the first quadrant and if, while ω is varied from 0 to $+\infty$ the locus of points corresponding to the imaginary roots 41 alternately cuts the lines $\eta = a_0$ and $\xi = a_1$, with the provision that the line $\eta = a_0$ is cut first and that the total number of points of intersection is $n-1$.

The advantage of this criterion lies in the simplicity of constructing the curve Γ_δ defined by equations 41; in the entire investigation being restricted to the real domain; and in the possibility of varying coefficients a_1 and a_0 without plotting a new curve. This will be demonstrated by the examples that follow.

Take a simple feedback system whose open-loop transfer function possesses three poles on the negative real axis

$$G(s) = \frac{K}{s(s+2)(s+3)} \tag{43}$$

The closed-loop transfer function is

$$\frac{C}{R}(s) = \frac{K}{s^3+5s^2+6s+K} \tag{44}$$

and the characteristic equation

$$s^3+5s^2+6s+K = 0 \tag{45}$$

which is identical to equation 29.

According to equation 15 the locus of points corresponding to the imaginary roots will be

$$\xi = \omega^2$$

$$\eta = 5\omega^2 \tag{46}$$

which is the straight line Γ_δ making an angle of $\tan^{-1} 5$ with the ξ axis. First, examine stability for the particular case $K=4$. The half-line Γ_δ, the straight lines $\eta = 4$, $\xi = 6$, and the point $M(6, 4)$ are shown in Fig. 9. Scale factors for the axes of ξ and η are chosen different for the sake of saving in space. It is seen that the point $M(6, 4)$ lies in the first quadrant, that the half-line Γ_δ alternately cuts the lines $\eta = 4$ and $\xi = 6$, starting with the line $\eta = 4$, and that the number of points of intersection is 2, i.e., exactly $n-1$. Thus for $K=4$ the system is stable.

Since K represents the ordinate of the point M, but is not present in equation 46, the variation of K results only in the point M being shifted along the line $\xi = 6$. The conditions for stability will evidently be fulfilled for all the points M between the ξ axis and Q_3. The first boundary value for K is zero and the second one will be determined from the condition requiring that the point M be on the line Γ_δ. The first one of the equations 46 gives $\omega^2 = 6$ and the second $\eta = K = 30$. Hence the system will be stable for all values of K in the interval $0 < K < 30$. The same result can also be obtained graphically by measuring the ordinate of the point Q_3.

If the coefficient $a_1 = 6$ is also allowed to vary, the point M may acquire any position in the $0\xi\eta$ plane. From the preceding discussion it follows, however, that the system under consideration will be stable only if the point M lies within the shaded, i.e., the stable region.

It is important to note that the locus Γ_δ associated with the second- and third-degree equations has always the form of a half-line. The stability of such systems may thus be examined without any calculation. The following example will show that higher-degree systems also present no particular difficulties.

Take the system described by the open-loop transfer function

$$G(s) = \frac{K(s+z_1)}{s(s+3)(s^2+2s+2)} \tag{47}$$

and examine the effect of varying parameters K and z_1 on the system's stability

The characteristic equation is

$$s^4+5s^3+8s^2+(6+K)s+Kz_1 = 0 \tag{48}$$

and the locus of points corresponding to the imaginary roots

$$\xi = 5\omega^2$$

$$\eta = 8\omega^2 - \omega^4 \tag{49}$$

The curve Γ_δ defined by equations 49 and the point $M(6+K, Kz_1)$ for $K=6$ and $z_1=1$ are shown in Fig. 10. It should be noted that for $n=4$ the curve Γ_δ is always part of a parabola whose axis is parallel to the axis of η. For an approximate plot it is therefore sufficient to find only the points A_1 and A_2.

By the use of the previously derived criterion of stability, it may easily be concluded that the system under consideration is stable for $K=6$ and $z_1=1$. The same case will occur whenever the point M falls within the stable region $0A_1$-A_20. Thus, assigning different values to K and z_1 and following the travelling of the point M, it is possible to conclude without any calculation not only whether a system is stable, but also whether the margin of stability is sufficient and what variations of K and/or z_1 contribute toward increasing or decreasing sta-

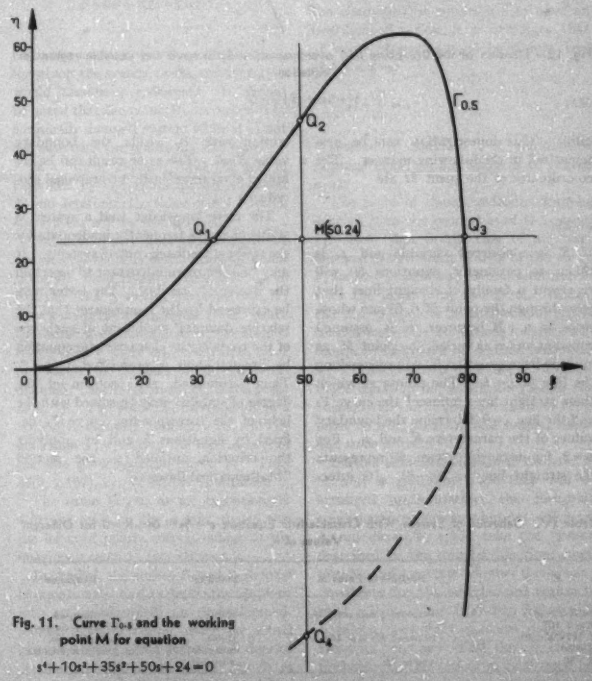


Fig. 11. Curve $\Gamma_{0.5}$ and the working point M for equation

$$s^4+10s^3+35s^2+50s+24 = 0$$

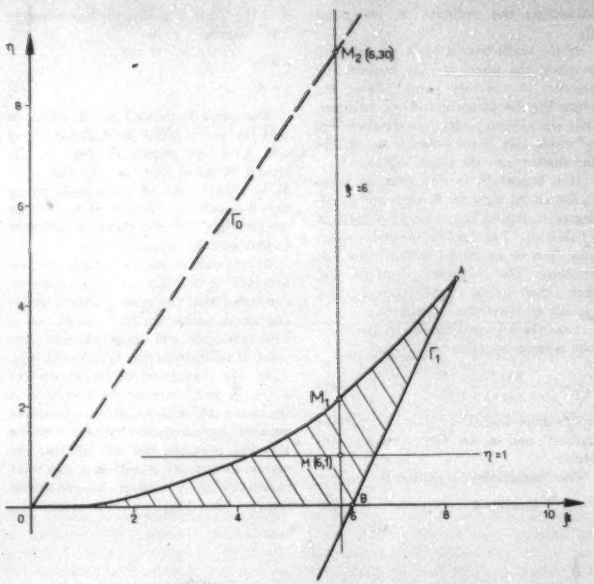


Fig. 12. Division of the $0\xi\eta$ plane into overdamped, underdamped and unstable region for equation

$$s^3+5s^2+\xi s+\eta=0$$

bility. This investigation can be systematized in the following manner. The co-ordinates of the point M are

$$\xi=6+K$$

$$\eta=Kz_1 \qquad (50)$$

If K is considered variable and z_1 is taken as parameter, equations 50 will represent a family of straight lines that pass through the point $M(6, 0)$ and whose slope is z_1. If however, K is assumed constant and z_1 is varied, the point M, as evident from equation 50 will move along the line $\xi=6+K$. The points at which these straight lines intersect the curve Γ_0 and the line $\eta=0$ determine the boundary values of the parameters K and z_1. For $z_1=2$, for instance, system 50 represents the straight line $\eta=2(\xi-6)$. Its intersection with Γ_0 yields the boundary value $K\approx 7$. The same result can be obtained even more easily by graphical procedure.

The mere knowledge that a system is stable or not is frequently inadequate for the study of feedback control systems. It may be even more important to ascertain the degree of stability. The latter may be expressed by the requirement that the relative damping coefficient of each one of the roots of the characteristic equation be greater than a prescribed value ζ. Thus formulated, the problem of the degree of stability may be solved with the help of the corresponding curve Γ_ζ defined by equations 5 and by applying the criterion outlined in the section "Mathematical Bases."

Table IV. Behavior of System With Characteristic Equation $s^3+5s^2+6s+K=0$ for Different Values of K

K	Position of Point M	Stability	Response
$0<K<2.107$	between ξ axis and M_1	stable	overdamped
$K=2.107$	M_1	stable	critically damped
$2.107<K<30$	between M_1 and M_2	stable	underdamped
$K=30$	M_2	unstable	purely sinusoidal
$K>30$	above M_2	unstable	

As an example take the system which is described by the characteristic equation

$$s^4+10s^3+35s^2+50s+24=0 \qquad (51)$$

and investigate whether the damping coefficient of all roots is greater than 0.5. As evident from equations 5 and 6 or equation 7, the curve $\Gamma_{0.5}$ is determined by the equations

$$\xi=35\omega_n-\omega_n^3$$

$$\eta=35\omega_n^2-10\omega_n^3 \qquad (52)$$

and is constructed in Fig. 11. It is seen that the point $M(50, 24)$ lies in the first quadrant, that the curve $\Gamma_{0.5}$ crosses the lines $\eta=24$ and $\xi=50$ four times alternately, intersecting, first, the line $\eta=24$ and that $m=4$ satisfies the condition expressed by inequality 14. The damping coefficients of all roots of equation 51 are hence greater than 0.5.

Overdamped Systems

For a feedback system to be overdamped, all roots of its characteristic equation must be real and negative. In other words, assuming $\omega_n>0$ the damping coefficient of every root must be greater than unity. (A sine function can never be greater than unity but it is obvious from equation 4 that for $\zeta>1$ the value s is real. One therefore assumes $\zeta>1$ to be the value corresponding to the real roots but it then loses its meaning as the sine of a real angle.) The problem of overdamped systems is obviously merely a special case of the conditions derived in the section "Mathematical Bases." The relevant curve, now, is Γ_1, i.e., the locus of points corresponding to the double real roots, the equations of which have already been derived under equation 23. Moreover, for $\zeta=1$, equation 4 yields $\theta=\pi/2$. Substituting this value in equation 14, one obtains $m=2(n-1)$ and the conditions for a system to be overdamped can be stated as follows:

The feedback control system described by the characteristic equation 2 will be overdamped if the point $M(a_1, a_0)$ lies in the first quadrant and if, as σ varies from 0 to $+\infty$, the locus of points corresponding to the double real roots and defined by equations 23 alternately cuts the straight lines $\eta=a_0$ and $\xi=a_1$, with the provision that the line $\eta=a_0$ is cut first and that the total number of points of intersection is $2(n-1)$.

Obviously, a system will be critically damped if the point $M(a_1, a_0)$ lies on the part of the curve Γ_1 defined by equations 23, which bounds the overdamped region. As an example, take the system

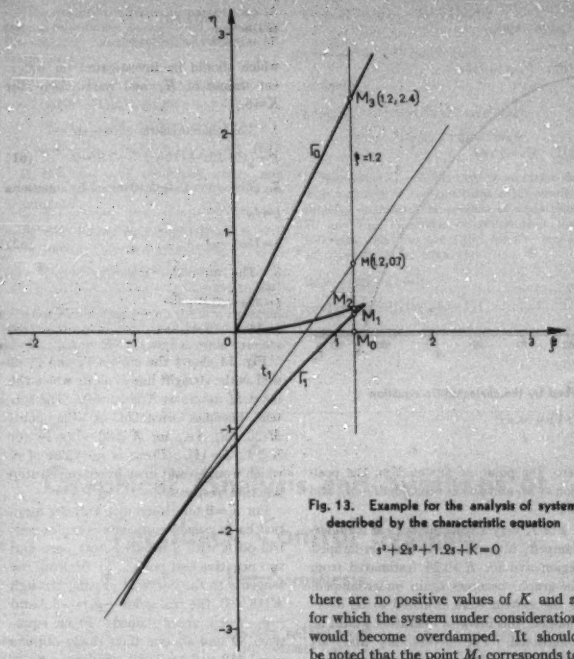


Fig. 13. Example for the analysis of system described by the characteristic equation

$$s^3+2s^2+1.2s+K=0$$

described by the open-loop transfer function

$$G(s)=\frac{K(s+z_1)}{s(s+4)(s^2+2s+4)} \tag{53}$$

and ascertain whether there exist any positive values of K and z_1 for which the system would be overdamped.

Since the characteristic equation is of the form 27, the locus of points corresponding to the double real roots, equations 23, will be determined by the equations 28.

The curve Γ_1 has been constructed with the help of Table II and is shown in Fig. 6. The point $M(9, 2)$ corresponds to an overdamped system since the curve Γ_1 has a total of $2(n-1)$ points of alternate intersection with the lines $\eta=2$, $\xi=9$, starting with the line $\eta=2$, and the point $M(9, 2)$ lies in the first quadrant. The entire area around the point $M(9, 2)$ enclosed by the curve Γ_1 represents the overdamped region. Therefore, the necessary condition for the point $M(16+K, Kz_1)$ to be within the overdamped region is

$$8<16+K<10 \tag{54}$$

This cannot be realized with a positive K regardless of the value of z_1. Hence there are no positive values of K and z_1 for which the system under consideration would become overdamped. It should be noted that the point M_1 corresponds to a critically damped system while it is not the case with the points M_2 and M_3.

As a further example, take the system described by the characteristic equation 29 and determine the value of K for which the system will be overdamped.

The locus of points corresponding to the double real roots is determined by equations 30. Once more, its plot is shown in Fig. 12. If K is varied, the point $M(6, K)$ moves along the straight line $\xi=6$. Since the latter goes through the overdamped region $0BA0$, there do exist values of K for which the system would be overdamped. Such, for instance, is the point $M(6, 1)$. The point M_1 corresponds to a critically damped system. It will be determined by seeking the intersection of the line $\xi=6$ and the curve Γ_1. Thus $M(6, 2.107)$, i.e., $K=2.107$ is obtained.

The same figure, at an exceptionally different scale on the η axis, shows also the locus of points corresponding to the imaginary roots Γ_0 (equations 46).

In general, the curves Γ_0 and Γ_1 defined by equations 41 and 23 divide the $0\xi\eta$ plane into an overdamped, an underdamped, and an unstable region, while the actual behavior of the given system will depend on the position of the point M. It is important to note that in the general case the underdamped region may be missing, implying that for no values of the coefficients a_1 and a_0 the system under consideration can become underdamped. The same applies for the overdamped region. Underdamped as well as overdamped systems may be studied by the use of the corresponding curves plotted on the same sheet since the point $M(a_1, a_0)$ is always common to all curves.

Table IV can be deduced directly from Fig. 12.

Necessary Conditions for Existence of Overdamped Region

In the preliminary studies it is sometimes necessary to know whether there is any possibility of making a system overdamped by moving the point $M(a_1, a_0)$. In other words, does the overdamped region exist? This question can obviously be answered by constructing the curve Γ_1 for the given characteristic equation 2. However, the following necessary but not sufficient conditions may provide useful information:

For a feedback system described by the characteristic equation 2 to have an overdamped region, it is necessary that all roots of equation 55 be simple, real and positive:

$$f'(-\sigma)=\frac{d\xi}{d\sigma}=0 \tag{55}$$

Instead of equation 2 it remains to examine equation 55, which is of degree $n-2$.

The proof of these conditions may be derived from the properties of the curves Γ_1. Note that in the case of an overdamped system each branch of the curve Γ_1 for $\sigma>0$ yields two points of intersection with the lines $\eta=a_0$ and $\xi=a_1$. Hence, the $2(n-1)$ points of intersection, representing one of the conditions for the overdamped region, can be realized only with $n-1$ branches, thus implying $n-2$ cusps. Since equation 55, which determines the cusps, is also of degree $n-2$, all its roots must be real, positive, and simple, which had to be proved.

Since a sufficient number of cusps is in no way a guarantee that they will be arranged appropriately, the foregoing conditions are necessary but not sufficient. As an example, again take the system described by the transfer function, equation 53, and ascertain whether there is any possibility for the overdamped region to exist. It is evident from Fig. 6 that such a region does exist (around point M). However, one may find that it is very probable for this region to exist even be-

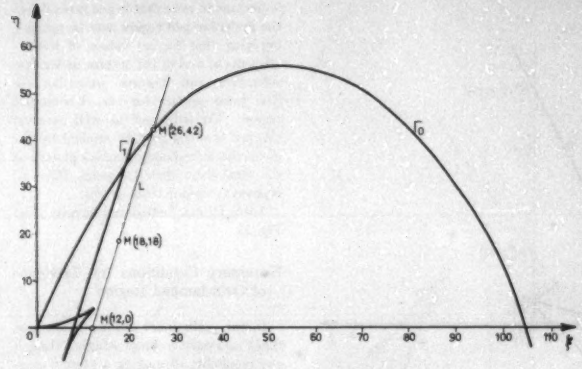


Fig. 14. Example for the analysis of system described by the characteristic equation

$$s^4+7s^3+15s^2+(12+K)s+3K=0$$

fore plotting the curve of Fig. 6. Differentiating twice equation 27 and setting $s=-\sigma$ yields $12\sigma^2-36\sigma+20=0$. The roots are 1 and 2. They are real, positive, and simple, which means that the necessary conditions are satisfied.

Examples for Analysis of Feedback Control Systems

As the first example, take the control system described by the transfer function

$$G(s)=\frac{K}{s(s^2+2s+1.2)} \tag{56}$$

and examine how it is affected by variation of the gain K. The characteristics for $K=0.7$ are of particular interest.

The following equations will be necessary for this investigation:

1. The characteristic equation

$$s^3+2s^2+1.2s+K=0 \tag{57}$$

2. The locus of points corresponding to the imaginary roots

$$\left.\begin{aligned}\xi&=\omega^2\\ \eta&=2\omega^2\end{aligned}\right\}\Gamma_0 \tag{58}$$

3. The locus of points corresponding to the double real roots

$$\left.\begin{aligned}\xi&=4\sigma-3\sigma^2\\ \eta&=2\sigma^2-2\sigma^3\end{aligned}\right\} \tag{59}$$

In addition to the curves Γ_0 and Γ_1, Fig. 13 shows also the straight line $\xi=1.2$ along which the point $M(1.2, K)$ moves as K is varied.

Since $K_v=K/1.2$ the ordinate of the point M is a measure of the velocity constant. As K is varied starting from zero, the point M moves from the position of M_0 and passes through M_1, M_2, and M_3.

For K very small, the system is underdamped; it then shifts to the overdamped region and for $K>0.25$ (estimated from the graph) becomes again underdamped. If the system were adjusted to be overdamped, this mode of operation would be conditional. The points M_1 and M_2 correspond to critically damped systems. At the point M_3 corresponding to $K_v=2$, the system is unstable. Hence K_v may vary between 0 and 2.

In general, this analysis is similar to one employed in a previous example and the results may be given in the form of Table IV. This analysis, however, may be further extended, as will be demonstrated by investigating the system for $K=0.7$. The corresponding point is $M(1.2, 0.7)$. There is only one tangent that can be drawn from this point to the curve Γ_1. Its slope is $\sigma=1.5$, showing that the closed-loop transfer function has only one real pole $p_1=-1.5$. The remaining two conjugate complex poles are determined from equations 39 and 40. Thus we obtain $\omega_n=0.683$ and $\zeta=0.366$. On account of equation 4 this yields $-p_{2,4}=-0.25\pm j0.635$. The velocity constant is $0.7/1.2=0.584$.

Since the pole $-p_1$ is approximately six times as far from the $j\omega$ axis as the control poles, its effect on the unit step function response is practically negligible and the system is thus reduced to the case of two control poles. The overshoot is less than 30%. As a further example, take the feedback control system described by the open-loop transfer function

$$G(s)=\frac{K(s+3)}{s(s+4)(s^2+35+3)} \tag{60}$$

which should be investigated for different values of K, and particularly for $K=6$.

1. The characteristic equation

$$s^4+7s^3+15s^2+(12+K)s+3K=0 \tag{61}$$

2. The curve Γ_0 is determined by equations

$$\left.\begin{aligned}\xi&=7\omega^2\\ \eta&=15\omega^2-\omega^4\end{aligned}\right\} \tag{62}$$

3. The curve Γ_1

$$\left.\begin{aligned}\xi&=30\sigma-21\sigma^2+4\sigma^3\\ \eta&=15\sigma^2-14\sigma^3+3\sigma^4\end{aligned}\right\} \tag{63}$$

Fig. 14 shows the curves Γ_0 and Γ_1 as well as the straight line L, along which the point M moves as K is varied. The system becomes unstable at the point $M(26, 42)$, i.e., for $K\geqq 26-12=14$ (or $K\geqq 42/3=14$). There is no value of K which would result in an overdamped step function response.

For $K=6$ the closed-loop transfer function has a pair of conjugate complex control poles with a negative real part and two negative real poles. By drawing the tangents to the curve Γ_1 passing through $M(18, 18)$, the real poles $-p_1=-2.2$ and $-p_3=-4.1$ are obtained. From equations 39 and 40 one then easily obtains $\omega_n=1.412$ and $\zeta=0.248$, yielding $-p_{2,4}=-0.35\pm j1.367$.

The effect of the poles $-p_1$ and $-p_3$ on the step function response is practically negligible since they are more than six times as far from the $j\omega$ axis as the control poles. Thus, for $K=6$ the system may be considered as a case of two control poles and one zero.

Appendix. Example Illustrating Procedure for Obtaining Equations 5 and 6

Let the following equation be given:

$$a_4s^4+a_3s^3+a_2s^2+a_1s+a_0=0 \tag{64}$$

After substituting variables ξ and η for a_1 and a_0, equation 64 becomes

$$a_4s^4+a_3s^3+a_2s^2+\xi s+\eta=0 \tag{65}$$

By introducing

$$s=-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2} \tag{66}$$

one obtains

$$\begin{aligned}&a_4(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})^4+\\ &\quad a_3(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})^3+\\ &\quad a_2(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})^2+\\ &\quad \xi(-\omega_n\zeta+j\omega_n\sqrt{1-\zeta^2})+\eta=0\end{aligned} \tag{67}$$

After raising the terms in parentheses to the corresponding powers and separating the real and imaginary parts, equation 67 takes the form

$$a_4(8\zeta^4-8\zeta^2+1)\omega_n^4+a_3(-4\zeta^3+3\zeta)\omega_n^3+a_2(2\zeta^2-1)\omega_n^2-\xi\omega_n\zeta+\eta+j\omega_n\sqrt{1-\zeta^2}[a_4(-8\zeta^2+4\zeta)\omega_n^3+a_3(4\zeta^2-1)\omega_n^2+a_2(-2\zeta)\omega_n+\xi]=0 \quad (68)$$

If the real and imaginary parts are equated to zero, a system of two equations is obtained

$$a_4(8\zeta^4-8\zeta^2+1)\omega_n^4+a_3(-4\zeta^3+3\zeta)\omega_n^2+a_2(2\zeta^2-1)-\xi\omega_n\zeta+\eta=0$$
$$a_4(-8\zeta^3+4\zeta)\omega_n^3+a_3(4\zeta^2-1)\omega_n^2+a_2(-2\zeta)\omega_n+\xi=0 \quad (69)$$

After eliminating ξ from the first equation of the system 69, the two equations should be solved for ξ and η, which results in

$$\xi=a_3(2\zeta)\omega_n+a_2(-4\zeta^2+1)\omega_n^2+a_4(8\zeta^2-4\zeta)\omega_n^3$$
$$\eta=-\omega_n^2[a_2(-1)+a_3(2\zeta)\omega_n+a_4(-4\zeta^2+1)\omega_n^2] \quad (70)$$

or finally

$$\xi=a_3\phi_3(\zeta)\omega_n+a_2\phi_2(\zeta)\omega_n^2+a_4\phi_4(\zeta)\omega_n^3$$
$$\eta=-\omega_n^2[a_2\phi_1(\zeta)+a_3\phi_3(\zeta)\omega_n+a_4\phi_2(\zeta)\omega_n^2] \quad (71)$$

Equations 71 are identical to equation 5. The meaning of the functions $\phi_i(\zeta)$ results directly from the comparison of equations 70 and 71. It should be noted that the following relations are valid for the terms in parentheses in equations 70:

$$2\zeta=-[2\zeta(-1)+0]$$
$$-4\zeta^2+1=-[2\zeta(2\zeta)+(-1)]$$
$$8\zeta^3-4\zeta=-[2\zeta(-4\zeta^2+1)+(2\zeta)] \quad (72)$$

or

$$\phi_3(\zeta)=-[2\zeta\phi_1(\zeta)+\phi_0(\zeta)]$$
$$\phi_4(\zeta)=-[2\zeta\phi_2(\zeta)+\phi_1(\zeta)]$$
$$\phi_5(\zeta)=-[2\zeta\phi_3(\zeta)+\phi_2(\zeta)]$$
$$\phi_6(\zeta)=-[2\zeta\phi_4-\phi_3(\zeta)+\phi_4-_2(\zeta)] \quad (73)$$

which shows how equation 6 is obtained.

References

1. Graphical Analysis of Control Systems. Walter R. Evans. *AIEE Transactions*, vol. 67, 1948, pp. 547–51.
2. The Extrapolation, Interpolation and Smoothing of Stationary Time Series (book). N. Wiener. John Wiley & Sons Inc., New York, N. Y., 1948.
3. Graphical Conditions for the Arguments of All Roots of an Algebraic Equation to Lie Between $\pi/2+\mu$ and $3\pi/2-\mu$ (in French). D. Mitrović. *Comptes Rendus*, Academie des Sciences, Paris, France, vol. 243, 1956, pp. 831–33.
4. Graphical Conditions for All Roots of an Algebraic Equation to Have Negative Real Parts (in French). D. Mitrović. *Ibid.*, vol. 240, 1955, pp. 1177–78.

Graphical Analysis and Synthesis of Feedback Control Systems

II—Synthesis

DUŠAN MITROVIĆ
NONMEMBER AIEE

WITH THE USE OF the method presented in Part I,[1] it is possible to develop several suitable procedures for the analysis and synthesis of feedback control systems. The purpose of all these procedures is to reduce the characteristic equation to a form where the unknown parameters would take part only in the co-ordinates of the working point $M(a_1, a_0)$. It is thereby achieved that changes in the values of parameters affect only the position of the working point and not the shape of the curve Γ_ζ. As a consequence, it is possible to estimate almost instantaneously all roots of the characteristic equation for any set of values of the unknown parameters and to select the latter without difficulty.

It is further necessary to adjust the synthesis procedure in such a way as to make the fullest use of the curves constructed once for all. Systems having characteristic equations up to and including the third degree, and also systems of a higher degree, which can be reduced by the cancellation method or some other ways to the third degree or less, can be solved without constructing the curves at all. If, however, curves are constructed in such cases also, this is done only in order to illustrate the general procedure.

Finally, during the synthesis, one should endeavour to realize, if possible, a system with only one or, at the utmost, two pairs of complex poles of the closed-loop transfer function. Under such conditions it is possible, on the basis of the position of the working point $M(a_1, a_0)$ to estimate all roots of the characteristic equation with practically no calculation.

The principal applications of the method proposed in Part I will be demonstrated by considering the system of Fig. 1. Since the proposed method is based on the study of the characteristic equation, the resulting conclusions are equally valid for more complicated systems. The closed-loop transfer function of the system of Fig. 1 is given by the equation

$$\frac{C}{R}(s)=\frac{p(s)}{f(s)}=\frac{G(s)}{1+G(s)} \quad (1)$$

where the open-loop transfer function is a rational function, i.e., a ratio of two polynomials in s:

$$G(s)=\frac{p(s)}{q(s)} \quad (2)$$

From equations 1 and 2 it follows immediately that

$$\frac{C}{R}(s)=\frac{p(s)}{p(s)+q(s)} \quad (3)$$

It is well known that the main difficulty arising in the design of feedback control systems by the use of the Laplace transformation is due to the fact that the poles of the closed-loop transfer function are equal to the zeros of the polynomial $p(s)+q(s)$. If $p(s)$ and $q(s)$ are given separately, the determination of zeros is not easy even in relatively simple cases. If, furthermore, it should be necessary to vary some of the parameters involved in $p(s)$ and/or $q(s)$, the procedure becomes too complicated to be suitable for practical use.

The proposed method solves the prob-

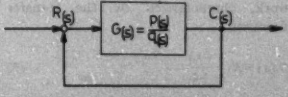


Fig. 1. Block diagram of a simple feedback control system

Paper 58-861, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted December 4, 1956; made available for printing April 22, 1958.

Dušan Mitrovic is with the University of Belgrade and the Boris Kidritch Institute, Belgrade, Yugoslavia.

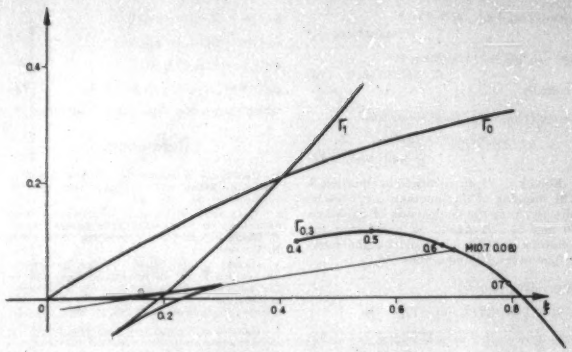


Fig. 2. Example for integral compensation

lem by adjusting $p(s)$ and $q(s)$ in such a way as to yield suitable or desired zeros of the polynomial $p(s)+q(s)$. The method permits directly the variation of two coefficients, which is its main advantage. By transformation or other indirect procedure to be explained later, it is often possible to select all parameters.

The discussion in Part I was based on the characteristic equation 2 (of Part I). The zeros of the polynomial $p(s)+q(s)$, however, are determined by seeking the roots of the equation $p(s)+q(s)=0$. Hence, instead of equation 2 in Part I, one may write

$$f(s)=p(s)+q(s)=a_n s^n+a_{n-1}s^{n-1}+\ldots+a_2 s^2+a_1 s+a_0=0 \quad (4)$$

It is seen that both $p(s)$ and $q(s)$ are present in the characteristic equation. By adjusting the roots of the equation while taking $p(s)$ and $q(s)$ into separate account, one in fact adjusts the poles and zeros of the closed-loop transfer function by choosing certain parameters of the open-loop transfer function. Since the discussion in Part I covers all that is necessary for application, the procedure will best be explained by examples.

Integral Compensation

One of the simplest ways of increasing the velocity constant without affecting the relative stability of the system is to introduce an integral compensation network, characterized by the transfer function

$$G_1(s)=K_1\frac{s+\lambda\delta}{s+\delta} \quad (5)$$

Generally, δ and $\lambda\delta$ should be small compared to the nonzero critical frequencies of the original open-loop transfer function. Furthermore $\lambda>1$.

To illustrate integral compensation by examining the system whose controlled part is described by the transfer function,

$$G_2(s)=\frac{K_2}{s(s^2+2s+1.2)} \quad (6)$$

The overshoot of the step function response is to be less than 30% and the velocity constant greater than 5/sec (per second). It is therefore necessary to increase K_2 while maintaining approximately the same relative stability. This can be achieved by integral compensation. The open-loop transfer function then becomes

$$G(s)=K\frac{s+\lambda\delta}{s(s^2+2s+1.2)(s+\delta)} \quad (7)$$

and the characteristic equation

$$s^4+(2+\delta)s^3+(1.2+2\delta)s^2+(K+1.2\delta)s+K\lambda\delta=0 \quad (8)$$

Since $\lambda\delta$ should be considerably less than unity, δ must be still less. Terms containing only δ may therefore be neglected, and equation 8 becomes

$$s^4+2s^3+1.2s^2+Ks+K\lambda\delta=0 \quad (9)$$

The corresponding curves Γ_0 and Γ_1 are determined by the equations

$$\left.\begin{aligned}\xi&=2\omega^2\\ \eta&=1.2\omega^2-\omega^4\end{aligned}\right\}\Gamma_0 \quad (10)$$

$$\left.\begin{aligned}\xi&=2.4\sigma-6\sigma^2+4\sigma^3\\ \eta&=1.2\sigma^2-4\sigma^3+3\sigma^4\end{aligned}\right\}\Gamma_1 \quad (11)$$

Both curves are shown in Fig. 2. It is now necessary to select such a position of the point $M(K, K\lambda\delta)$ as to ensure that the specified conditions will be fulfilled. Suppose the point $M(a_1, a_0)$ is selected. The following relations then hold:

$$K=a_1 \quad (12)$$

$$K\lambda\delta=a_0 \quad (13)$$

$$K_v=\frac{K\lambda}{1.2} \quad (14)$$

To select the correct position of the point M, consult Fig. 2. It is seen that for all points in the neighborhood and to the right of the point M_1, the closed-loop transfer function has two conjugate complex and two negative real poles. The smaller of the two real poles and the zero of the closed-loop transfer function form a dipole whose effect may be neglected. The other real pole is large and its effect is also negligible. The step function response will therefore be governed by only one pair of poles and the value of the damping coefficient corresponding to a 30% overshoot will be approximately $\zeta=0.3$. One should construct therefore the curve Γ_ζ for $\zeta=0.3$, and select the working point to the right of the point M_1. The specification on the overshoot would thus be automatically satisfied. The equations of the curve are

$$\left.\begin{aligned}\xi&=0.72\omega_n+1.28\omega_n^2-0.948\omega_n^3\\ \eta&=1.2\omega_n^3-1.2\omega_n^3-0.64\omega_n^4\end{aligned}\right\}\Gamma_{0.3} \quad (15)$$

and its shape as shown in Fig. 2. Equations 12, 13, and 14 will then yield such values of K, λ and δ, as required by the specified velocity constant. Obviously the choice of the working point is by no means critical, permitting some other specifications to be taken into account as well.

If, for instance, one selects the same value of K as in the corresponding example of Part I (this value resulted in $\zeta=0.366$), the point $M(a_1, a_0)$ will be fixed automatically. According to Fig. 2, this point is $M(0.7, 0.08)$. Equations 12, 13, and 14 then give $\lambda=8.57$ and $\delta=0.0133$.

The most accurate procedure for obtaining the pair of control poles is the following: The damping coefficient is already known ($\zeta=0.3$) and ω_n can be determined by interpolation along the curve $\Gamma_{0.3}$. Thus one obtains $\omega_n=0.617$; the control poles are $-0.185\pm j0.589$. The real poles may be determined by drawing the tangents through M. The smaller pole is -0.14 while the larger one cannot be determined accurately by drawing the tangent. The effect of the latter on the step function response, however, is negligible and an accurate determination is not necessary. Nevertheless, if this be desirable, the more accurate results will be obtained by the use of equation 39 of Part I. Thus the value -1.5 is obtained. The two real poles could as well have been determined by the use of equations 39 and 40 of Part I.

If it is required to control the amplitude of the long-time constant tail of the transient response, the dipole separation of the closed-loop transfer function must be specified. This actually signifies that the ratio of the respective pole and zero must not exceed a given value slightly greater than unity. This problem can easily be solved by the proposed method. From the co-ordinates of the working point $M(K, K\lambda\delta)$ it follows immediately that the slope of the radius vector is equal to the zero of the closed-loop transfer function (with a changed sign). Furthermore, the slope of the tangent drawn from M to Γ_1 determines the pole of the dipole. Therefore the radius vector and the tangent which is approximately parallel to it determine the dipole of the closed-loop transfer function. To satisfy the specified dipole separation, the point M should be chosen in such a way that the slopes of the radius vector and the tangent have the specified ratio. Which of the many such points will finally be chosen depends on other requirements.

It is seen from the foregoing discussion that the choice of the working point may be governed by very different conditions. This is not cumbersome since the proposed method permits simultaneous estimate of all poles and zeros.

In the example given in Fig. 2 the curve Γ_1 has two cusps. It may occur, however, that the curve Γ_1 has no cusps (or a reduced number of them). This in no way affects the procedure, except that the curve Γ_1 has a different shape.

Lead Compensation

This compensation is effective in increasing the velocity constant when all poles of the uncompensated open-loop transfer function are real. Therefore it remains to examine the system whose fixed part is described by the transfer function

$$G_2(s)=\frac{K_1}{s(s+4)} \tag{16}$$

and impose the following requirements: bandwidth less than 80 radians/sec; velocity error constant 250/sec; damping coefficient of the control poles approximately 0.5% and the overshoot of the step function response less than 30%.

The system described by the transfer function $G_2(s)$ cannot meet these requirements. Let us therefore try to introduce a lead compensation network. The open-loop transfer function then becomes

$$G(s)=K\frac{s+\alpha}{s(s+4)(s+\beta)} \tag{17}$$

and the task is to determine such values of α and β as to satisfy the specifications.

Since the value of β is not critical, its choice may be based on the following reasoning: For $\zeta=0.5$ the bandwidth does not exceed $2\omega_n$. This implies that the specification on bandwidth will be met by choosing $\omega_n=40$ for the control poles of the closed-loop transfer function. Also, the third pole of the closed-loop transfer function should be at least six times as large as the real part in order that its effect might be negligible. This condition would certainly be satisfied if $\beta=6.40=240$ is chosen. With $\beta=236$, which is more convenient for handling, the equation of the system becomes

$$s^3+240s^2+(944+K)s+K\alpha=0 \tag{18}$$

Substitution $s=200p$ results in the transformed equation

$$p^3+1.2p^2+(236+0.25K)\times10^{-4}p+0.125K\alpha\times10^{-6}=0 \tag{19}$$

The corresponding curves are

$$\left.\begin{aligned}\xi&=\omega_1^3\\ \eta&=1.2\omega_1^2\end{aligned}\right\}\Gamma_{1i} \tag{20}$$

$$\left.\begin{aligned}\xi&=2.4\sigma_1-3\sigma_1^3\\ \eta&=1.2\sigma_1^2-2\sigma_1^3\end{aligned}\right\}\Gamma_{1i} \tag{21}$$

$$\left.\begin{aligned}\xi&=1.2\omega_{n1}\\ \eta&=1.2\omega_{n1}^2-\omega_{n1}^3\end{aligned}\right\}\Gamma_{0.5i} \tag{22}$$

All three curves are constructed in Fig 3. Points on the curves Γ_{1i} and $\Gamma_{0.5i}$ are labelled by the corresponding values of σ_1, ω_{n1}. If the specification concerning the damping coefficient is to be satisfied, the working point must obviously lie on the curve $\Gamma_{0.5i}$. From equation 20 of Part I it follows $\omega_n=200\,\omega_{n1}$, i.e., numbers on the curve multiplied by 200 give the values for ω_n.

It has already been established that the specification on bandwidth will be satisfied if $\omega_n=40$ is chosen ($\omega_{n1}=0.2$). Thus, by selecting the working point on the curve $\Gamma_{0.5i}$ at $\omega_{n1}=0.2$, all specifications except the one concerning the velocity constant will be met. It remains still to check this remaining specification.

The working point is $M(0.24,\ 0.04)$, which, on account of $\zeta=0.5$ and $\omega_n=40$, corresponds to the control poles $-20\pm j34.6$. The third real pole of the closed-loop transfer function is determined by drawing the tangent to Γ_{1i}. Thus is obtained $-p_3=-200$, implying that the effect of this pole on the transient response is negligible. Also, by equating the co-ordinates of the point M to a_1 and a_0 of 19, $K=8656$ and $\alpha=37$ are obtained. The poles and zeros of the closed-loop transfer function being known, the velocity error constant can be determined from the well-known expression

$$\frac{1}{K_v}=\sum\frac{1}{p_k}-\sum\frac{1}{z_k} \tag{23}$$

where $-p_k$ and $-z_k$ are respectively poles

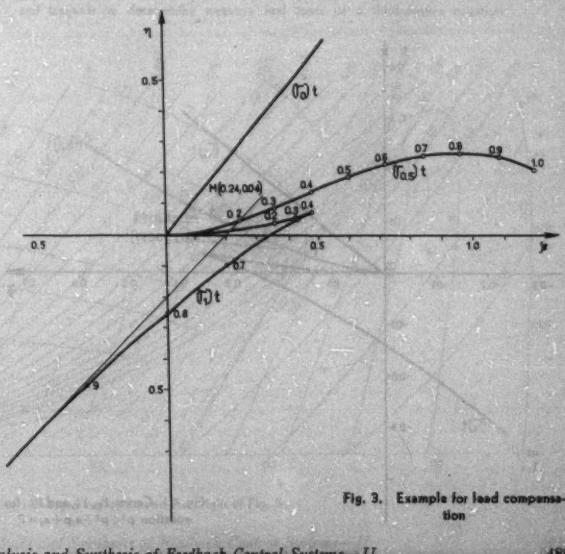


Fig. 3. Example for lead compensation

and zeros of the closed-loop transfer function. Thus $K_v=340$/sec is obtained, showing that the specification on the velocity constant is also satisfied. If this were not the case, it could easily be concluded from equation 23, in which direction the point M should be moved along the curve $\Gamma_{0\cdot 3\xi}$. It would then be necessary to give attention to a possible change in the bandwidth.

Following the foregoing discussion, the lead compensation network is described by the transfer function

$$G_1(s)=\frac{s+37}{s+236} \tag{24}$$

The following closed-loop transfer function is also obtained:

$$\frac{C}{R}(s)=8{,}656\times\frac{s+37}{(s+20+j34.6)(s+20-j34.6)(s+200)} \tag{25}$$

and the system characteristics can be checked by plotting the exact curve of the step function response. Since all poles of the closed-loop transfer function are known, the inverse Laplace transformation can easily be performed with the help of Heaviside expansion theorem.

The design procedure might be even more straightforward (like that of the next section). Namely, without choosing the value of β, the characteristic equation can be derived from equation 17 and then transformed by substituting $s=(4+\beta)p$.

Cancellation Compensation

If the open-loop transfer function contains conjugate complex poles, the lead compensation is often ineffective and one has to resort to a compensation transfer function with two zeros and two poles. The zeros can then be chosen so as to cancel exactly the poles of the uncompensated transfer function. Thus the problem is actually solved by choosing the zeros of the compensation transfer function and by seeking for poles that will meet the specifications. The cancellation compensation can also be handled by the proposed method. This will be illustrated in the following example with the use of the uncompensated transfer function described by

$$G_2(s)=\frac{K}{s(s^2+2s+1.2)} \tag{26}$$

After the usual application of the cancellation method the open-loop transfer function becomes

$$G(s)=K\frac{s^2+2s+1.2}{s^3+\lambda_1 s+\lambda_0}\times\frac{1}{s(s^2+2s+1.2)}=\frac{K}{s(s^3+\lambda_1 s+\lambda_0)} \tag{27}$$

The corresponding characteristic equation is

$$s^3+\lambda_1 s^2+\lambda_0 s+K=0 \tag{28}$$

The solution of the problem consists in determining K, λ_1 and λ_0 in such a way that the specifications are met. One of the procedures is to choose a value for λ_1, starting from the requirement that the velocity constant must be increased by a factor of 10. This automatically entails an increase of λ_1 by the same factor of 10, i.e., $\lambda_1=20$. It now remains to choose the working point in accordance with other specifications. The only specification left is that overshoot must be less than 30%. This implies, under the assumption that the transient response will be governed by only one pair of poles, that the damping coefficient should be approximately 0.3.

For the sake of greater accuracy, first, insert in equation 28 the substitution $s=20\rho$. The transformed equation takes the form

$$\rho^3+\rho^2+\frac{\lambda_0}{400}\rho+\frac{K}{8{,}000}=0 \tag{29}$$

The equations of the curves $\Gamma_{0\xi}$, $\Gamma_{1\xi}$ are

$$\left.\begin{aligned}\xi&=\omega_\xi{}^3\\ \eta&=\omega_\xi{}^3\end{aligned}\right\}\Gamma_{0\xi}\qquad\left.\begin{aligned}\xi&=2\sigma_\xi-3\sigma_\xi{}^3\\ \eta&=\sigma_\xi{}^3-2\sigma_\xi{}^3\end{aligned}\right\}\Gamma_{1\xi} \tag{30}$$

The equation of the curve $\Gamma_{0\cdot 3\xi}$ are

$$\left.\begin{aligned}\xi&=0.6\omega_{n\xi}+0.64\omega_{n\xi}{}^3\\ \eta&=\omega_{n\xi}{}^3-0.6\omega_{n\xi}{}^2\end{aligned}\right\}\Gamma_{0\cdot 3\xi}$$

All three curves are plotted in Fig. 4. Since the curves $\Gamma_{\xi t}$ will be of use in determining the roots in many other cases, they are plotted separately in Figs. 5 and 6 to a larger scale, together with the tangents. With the help of Figs. 5 and 6 and the substitution $s=\alpha_2\rho/a_1$, all third-order equations can be solved without calculations.

If the working point on the curve $\Gamma_{0\cdot 3\xi}$ is selected at $\omega_{n\xi}=0.4$, i.e., if $M(0.34, 0.122)$ is chosen, the comparison with 29 yields $\lambda_0=136$ and $K=976$. The control poles of the closed-loop transfer function are $-2.4\pm j7.62$; by drawing the tangents or, still better, by the use of equation 39 of Part I, the real pole is obtained as -15.2. In accordance with the assumption made, the effect of this pole is negligible. The velocity error constant is $K_v=K/\lambda_0=7.17>7$/sec. All specifications now being met, the final values $K=976$, $\lambda_1=20$ and $\lambda_0=136$ may be selected.

For the compensation transfer function to be realizable as an R-C (resistance-capacitance) network, all its poles must be negative, real and simple. For $\lambda_1=20$ and $\lambda_0=136$, however, these poles are conjugate complex. The compensation transfer function could be made realizable as an R-C network by moving the working point M along the curve $\Gamma_{0\cdot 3\xi}$ in the direction of the decreasing values of $\omega_{n\xi}$, but the velocity error constant then falls below 7/sec. This implies that for the requirement

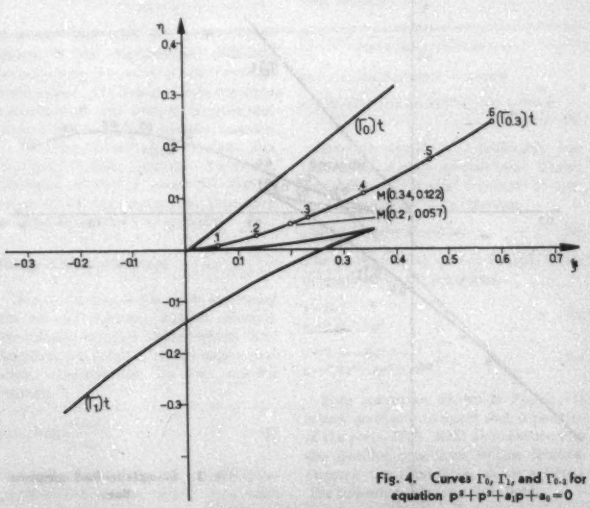


Fig. 4. Curves Γ_0, Γ_1, and $\Gamma_{0\cdot 3}$ for equation $p^3+p^2+a_1p+a_0=0$

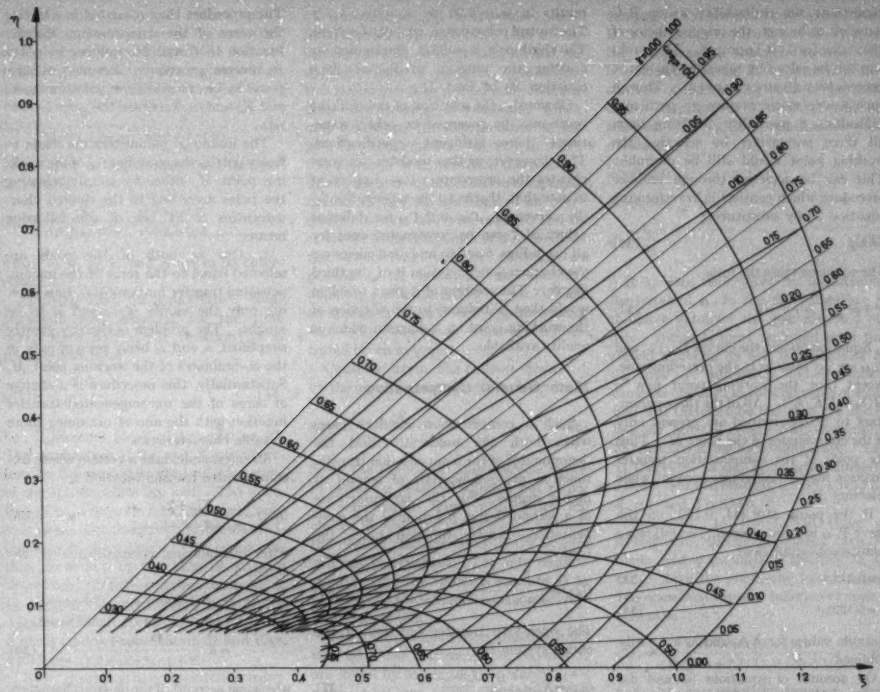


Fig. 5. Curves Γ_ξ for different values ξ and tangents for determining negative real roots of a third-degree equation

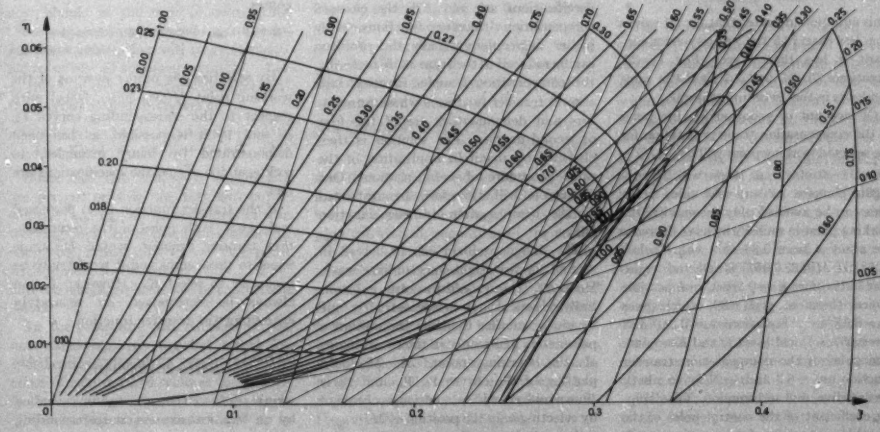


Fig. 6. Enlarged region around the origin of Fig. 5

concerning the realizability as an R-C network to be met, the original choice of the value for λ_1 is improper. This problem can be solved by repeating the entire process for a greater value of λ_1. Though such a step would involve no particular difficulties, a possibility of acting upon all three parameters by selecting the working point would still be desirable. This can be achieved through another procedure which consists in transforming equation 28 by substituting

$$s=\lambda_1 p \tag{31}$$

The equation takes the form

$$p^3+p^2+\frac{\lambda_0}{\lambda_1^2}p+\frac{K}{\lambda_1^3}=0 \tag{32}$$

Equation 30 and the curves Γ_{0l}, Γ_{1l}, and $\Gamma_{0\text{-}1l}$ are the same as in the preceding case, except that the working point now is $M(\lambda_0/\lambda_1^2, K/\lambda_1^3)$. All of the three parameters to be determined are present only in the co-ordinates of the point M. Thus the poles of the compensation transfer function may be taken into account while selecting M.

If the point $M(0.344, 0.122)$ on the curve $\Gamma_{0\text{-}1l}$ is again selected, the following relations obviously hold

$$\lambda_0=0.34\lambda_1^2 \tag{33}$$

$$K=0.122\lambda_1^3 \tag{34}$$

Suitable values for $K_1\lambda_1$ and λ_0 are easily obtained.

On account of equations 33 and 34, equation 27 yields

$$K_v=\frac{0.122}{0.34}\lambda_1=0.359\lambda_1 \tag{35}$$

This equation indicates that for all values of λ_1 less than 19.5 the velocity error constant is less than 7/sec. For $\lambda_1=20$, equations 33, 34, and 35 yield the same results as in the preceding example.

On account of equation 31, the poles of the compensation transfer function for the selected working point will be $-\lambda_1/2\pm\sqrt{\lambda_1^2/4-0.34\lambda_1^2}$, i.e., they will be conjugate complex for any real value of λ_1. This can be avoided only by moving the working point in such a way as to decrease the abscissa from 0.34 to a value below 0.25. If $M(0.2, 0.057)$ is selected, equation 35 becomes $K_v=0.285\,\lambda_1$. A possible choice, therefore, is $\lambda_1=30$, which gives $K_v=8.55$/sec. Equations $\lambda_0=0.2\lambda_1^2$ and $K=0.057\,\lambda_1^3$ yield $\lambda_0=180$ and $K=1{,}540$. The poles of the compensation transfer function are -8.3 and -21.7, i.e., both are negative, real and simple. The damping coefficient of the control poles of the closed-loop transfer function is $\zeta=0.3$, while interpolation along the curve $\Gamma_{0\text{-}1l}$ results in $\omega_{n1}=0.26$ or $\omega_n=30\omega_{n1}=7.8$. The control poles, hence, are $-2.34\pm j7.44$. The third pole is -25.3, determined by drawing the tangent to Γ_{1l} or from equation 39 of Part I.

Obviously, the selection of the working point may be governed by other, sometimes more stringent, specifications. This, however, neither modifies nor complicates the procedure. It is important to note that, thanks to the substitution 31 the curves Γ_{0l}, Γ_{1l}, and $\Gamma_{\zeta l}$ for different values of ζ can be constructed once for all (as in Figs. 5 and 6) and used whenever the characteristic equation is of the third degree. The solution of a given problem would then be reduced to the selection of the working point on a diagram which is readily available.

More General Compensation

Such a compensation involves cases when both the numerator and the denominator of the compensation transfer function are polynomials of second or higher degree. For the particular case of a second degree polynomial the compensation transfer function takes the form

$$G_1(s)=\frac{s^2+\nu_1 s+\nu_0}{s^2+\lambda_1 s+\lambda_0} \tag{36}$$

The open-loop transfer function of the entire system is

$$G(s)=KG_2(s)\frac{s^2+\nu_1 s+\nu_0}{s^2+\lambda_1 s+\lambda_0} \tag{37}$$

The task is to determine the coefficients ν_1, ν_0, λ_1, λ_0, and K in such a way that the specifications are met. In the case of compensation transfer functions with higher degree polynomials the problem can be reduced to the form 36 by factoring the compensation transfer function into several transfer functions whose numerators and denominators are of the first or second degree. The solution is then obtained by manifold application of the same procedure of compensation. Our discussion will therefore be restricted to the compensation transfer function 36.

The design problem can be solved in several, at first sight very unlike, ways. However, the procedures are substantially identical. It is necessary by one means or another to make the unknown parameters figure only in the co-ordinates of the working point M. Manifold plotting of the curves Γ_0, Γ_1, and Γ_ζ is thus avoided and the solution is obtained by selecting only the position of M.

A particular case has already been considered under cancellation compensation. The procedure then consisted in selecting the zeros of the compensation transfer function $G_1(s)$ and determining its poles. A reverse procedure, however, is also possible, i.e., to select the poles (or λ_1, λ_0, and K) and to determine the zeros (ν_1 and ν_0).

The unknown parameters are made to figure only in the co-ordinates of the working point M, either by simply selecting the poles according to the desired characteristics or by one of the following means:

1. One or both of the poles are selected equal to the zeros of the uncompensated transfer funtion $G_2(s)$, thus leaving only the values of ν_1 and ν_0 to be sought. The problem is thereby greatly simplified, ν_1 and ν_0 being present only in the co-ordinates of the working point M. Substantially, this procedure is a change of zeros of the uncompensated transfer function with the aim of obtaining more suitable characteristics.

As an example, take a system whose uncompensated transfer function is

$$G_2(s)=\frac{K(s+2)}{s(s+3)(s^2+2s+2)} \tag{38}$$

After introducing compensation it becomes

$$G_T(s)=\frac{K(s+2)}{s(s+3)(s^2+2s+2)}\times\frac{s^2+\nu_1 s+\nu_0}{(s+2)(s+\beta)}$$
$$=K\frac{s^2+\nu_1 s+\nu_0}{s(s+3)(s^2+2s+2)(s+\beta)} \tag{39}$$

where, on account of the desired characteristics K and β should be selected first. If, for instance, $K=500$ and $\beta=10$ are chosen, the characteristic equation becomes

$$s^5+15s^4+58s^3+586s^2+(60+500\nu_1)s+500\nu_0=0 \tag{40}$$

By selecting the proper position of the working point $M(60+500\nu_1, 500\nu_0)$ with respect to the corresponding curves Γ_0, Γ_1 and Γ_ζ it is possible, as has been demonstrated by many examples, to seek ν_1 and ν_0 so that the specifications are met.

If the transfer function $G_2(s)$ possessed real zeros, both poles of the compensation transfer function would be made equal to these zeros and it would only be necessary to select K. However, K can always be determined by separately examining the transfer function

$$G_T=\frac{K}{s(s+3)(s^2+2s+2)(s+\beta)} \tag{41}$$

where $s+\beta$ is either selected or cancelled by an identical expression corresponding to the second zero of the uncompensated transfer function.

2. Another procedure reduces substantially to an extension of the cancellation compensation. This will be demonstrated by the example of a system whose uncompensated transfer function is

$$G_x(s)=\frac{K}{s(s+1)(s+3)}=\frac{K}{s(s^2+4s+3)} \quad (42)$$

The corresponding compensated transfer function is

$$G(s)=\frac{K}{s(s^2+4s+3)}\times\frac{s^2+\nu_1 s+\nu_0}{s^2+\lambda_1 s+\lambda_0} \quad (43)$$

where ν_1, ν_0, λ_1, λ_0, and K are to be determined.

Applying first the cancellation compensation and assuming for the time being $\nu_1=4$ and $\nu_0=3$, equation 43 is transformed into

$$G(s)=\frac{K}{s(s^2+\lambda_1 s+\lambda_0)} \quad (44)$$

i.e., it becomes identical to equation 27. The parameters λ_1, λ_0, and K can therefore be determined so that at least some of the specifications are met. The selection of parameters serves sometimes only as a means of making possible the succeeding stage in the solution of the compensation problem. In the case of a third degree characteristic equation it is possible to perform this preliminary part of the solution without constructing any curves, if use is made of substitution 31 and Figs. 5 and 6.

Now, return to the compensation transfer function of the form 43 and substitute values found for λ_1, λ_0, and K. To obtain the latter, it was temporarily assumed that $\nu_1=4$ and $\nu_0=3$. However, nothing now stands in the way of considering these parameters as unknown and of seeking those values that will make it possible to satisfy all specifications fully. Let us therefore write the characteristic equation corresponding to system 43.

$$s^5+(4+\lambda_1)s^4+(3+4\lambda_1+\lambda_0)s^3+(3\lambda_1+4\lambda_0+K)s^2+(3\lambda_0+K\nu_1)s+K\nu_0=0 \quad (45)$$

Since the two unknown parameters are present only in the last two coefficients, i.e., in the co-ordinates of the working point M, it now remains only to select the proper position of this point with respect to the curves Γ_0, Γ_1, and Γ_ξ, and the design problem will be completely solved.

In the procedure just outlined it was not stated explicitly to what extent the specifications should be met in the first stage of solution and what is to be done to secure success in the second stage. The most logical procedure would be to try in the first stage to satisfy the specifications as much as possible as though the second stage would never take place. Such a procedure may often lead to the solution of the problem and should be applied whenever possible. One objection is that two pairs of conjugate complex poles are usually encountered, which may complicate the interpretation of the pole-zero configuration. A way to avoid this difficulty is to approach the problem during the stage of cancellation compensation as though the synthesis of an overdamped system has been required and to endeavor to meet as much of the specification concerning velocity constant as possible. If, however, the specified damping coefficient is large ($\zeta=0.7$), such an approach is not effective in increasing the velocity constant. This will be illustrated by an example.

Consider again the system described by the uncompensated transfer function

$$G_x(s)=\frac{K}{s(s^2+2s+1.2)} \quad (46)$$

Since the purpose of this example is not to present a practical solution but rather to demonstrate what possibilities are offered by applying the proposed method, the desired characteristics of the system will not be specified. The only requirement is that the velocity constant is to be increased while maintaining a relatively small bandwidth.

Take the compensation transfer function 36. The parameters λ_1, λ_0, and K are found by assuming temporarily $\nu_1=2$ and $\nu_0=1.2$, and applying compensation cancellation according to equation 27. This results in the transformed characteristic equation 30 and the curves shown in Figs. 4, 5, and 6. In the first stage of solution it is necessary only to select the position of the working point M.

It should be noted that the characteristic equation in the case of cancellation compensation, in fact, is

$$(s^3+\lambda_1 s^2+\lambda_0 s)(s^2+2s+1.2)+K(s^2+2s+1.2)=0 \quad (47)$$

and that it is reduced to 28 by a very formal cancellation. Hence the closed-loop transfer function of the actual physical system, regardless of the coefficients λ_1, λ_0, and K always possesses a pair of conjugate complex poles equal to the cancelled poles of the transfer function $G_x(s)$. If the system under consideration is to be governed by only one pair of conjugate complex control poles, the working point should be selected in such a way that all roots of equation 28 are real and negative. Such a system would theoretically be overdamped.

It now remains to select from among the infinite number of points in the overdamped region the most appropriate one. In the preceding section it was shown that the compensation transfer function will be realizable as an R-C network if the abscissa of the working point is less than 0.25. Therefore $M(0.24,\ 0.01)$ will be selected. The following relations are then obtained.

$$\lambda_0=0.24\lambda_1^2$$

$$K=0.01\lambda_1^2$$

$$K_v=\frac{\lambda_1}{24} \quad (48)$$

If λ_1 were determined by introducing the required K_v in the third equation, a too large value for the gain K would result. Hence λ_1 must be found from the second equation 48 and the choice of K as large as admissible. The velocity constant obtained in the first stage of solution will therefore be less than the specified one, but the situation will eventually be remedied in the succeeding stage. Let us therefore choose K as approximately 9,000 which will result in $\lambda_1=96$, $\lambda_0=2212$, $K_v=4$, and $K=8{,}850$.

Proceeding to the second stage of solving the specified problem assume that ν_1 and ν_0 are variable and that λ_1, λ_0, and K have the previously found values. The open-loop transfer function of such a system is

$$G(s)=\frac{8{,}850(s^2+\nu_1 s+\nu_0)}{s(s^2+96s+2{,}212)(s^2+2s+1.2)} \quad (49)$$

and the characteristic equation, after substituting $s=10\rho$, becomes

$$\rho^5+9.8\rho^4+24.05\rho^3+13.39\rho^2+(0.2653+0.885\nu_1)\rho+0.0885\nu_0=0 \quad (50)$$

The corresponding curve $\Gamma_{1\xi}$ and one section of $\Gamma_{0\xi}$ are drawn in Fig. 7. In the further solution of the problem only the small part of the stable region near the origin is pertinent. It is shown enlarged in Fig. 8 (scale factors for the two axes are different). A section of the curve $\Gamma_{0\text{-}7\xi}$ which may also be of interest in the succeeding study of this example is shown in the same figure. The point $M(2.035,\ 0.1061)$ corresponds to the system compensated by cancellation ($\nu_1=2$, $\nu_0=1.2$). For arbitrary values of ν_1 and ν_0, however, the point $M(0.2653+0.885\nu_1,\ 0.0885\nu_0)$ may acquire any position in the first quadrant to the right of the line $\xi=0.2653$. The selection of its position and consequently the selection of the parameters ν_1 and ν_0 is governed by the desired system characteristics.

All points between $\Gamma_{0\xi}$ and ξ axis belong to the stable region while those between the curve $\Gamma_{1\xi}$ and ξ axis belong to the overdamped region. The latter is shown in

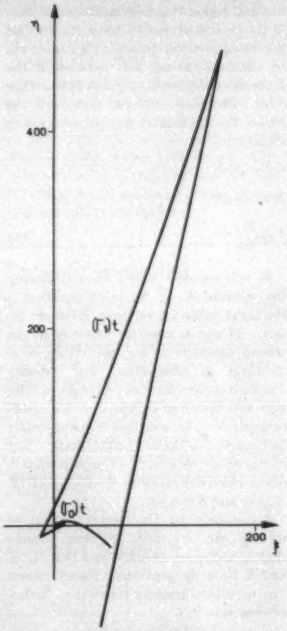


Fig. 7. Curve Γ_1 and part of the curve Γ_0 corresponding to the equation $p^5+9.8p^4+24.05p^3+13.39p^2+(0.2653+0.885\nu_1)p+0.0885\nu_0=0$

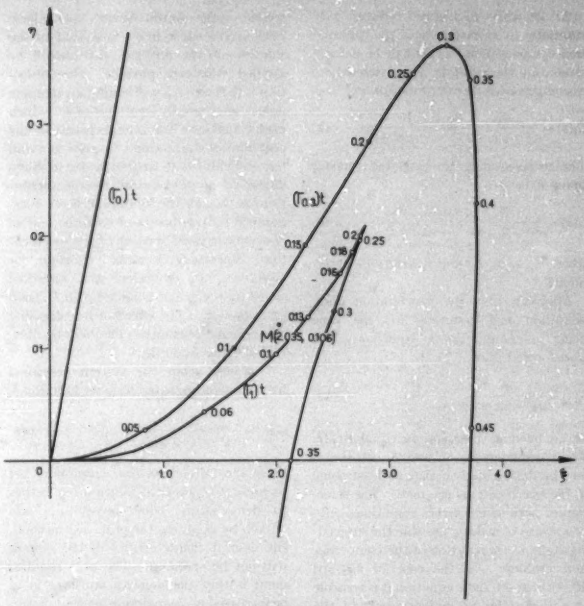


Fig. 8. Enlarged region around the origin of co-ordinates for the example of Fig. 7

Fig. 8 in its entirety. The points M between Γ_{1t} and $\Gamma_{0\cdot7t}$ are associated with systems whose pair of control poles has a damping coefficient between 0.7 and 1.0. Between the curves $\Gamma_{0\cdot7t}$ and Γ_{0t}, ζ is less than 0.7. By varying the values of ν_1 and ν_0 an overdamped, a critically damped, or an underdamped system may be realized.

One of the main problems to be solved now, by the selection of a new position of the working point, is to satisfy the specification concerning the velocity constant. The larger the specified damping coefficient of the control poles of the closed-loop transfer function the more difficult this problem is to solve. The curve $\Gamma_{0\cdot7t}$ in Fig. 8 is drawn, assuming $\zeta=0.7$. Since the velocity constant is proportional to the ordinate of the working point M, the latter should be selected as close to the maximum of the curve $\Gamma_{0\cdot7t}$ as possible. Therefore choose the point $\omega_{nt}=0.3$ whose co-ordinates are accurately known from the construction of the curve; it is $M(3.63,\ 0.372)$. The parameters $\nu_1=3.8$, $\nu_0=4.2$, and $K_v=14/\text{-sec}$ are thereby also determined. A characteristic to be noticed with this type of compensation is a relatively low value of velocity constant obtained for very large values of gain. The values found for ν_1 and ν_0 result in complex zeros of the compensation transfer function. Should it be desirable that these zeros are real, the maximum realizable velocity constant would be somewhat less (approximately 12.5/sec).

From Fig. 7 it is seen that the two large-magnitude negative poles of the closed-loop transfer function are very slightly changed by moving the point M. Hence they can be determined even before the point is selected. Anyway, their effect is negligible. The third, smaller pole, can be determined from the tangent to the curve Γ_{1t} in Fig. 8. Since, in addition, the pair of control poles for all points on the curve $\Gamma_{0\cdot7t}$ is known, all poles and zeros of the closed-loop transfer function can be estimated straightforward while the point M is being selected. For the particular case of $M(3.63,\ 0.372)$ the closed-loop transfer function can be written approximately in the form

$$\frac{C}{R}(s) = \frac{8{,}850(s^2+3.8s+4.2)}{(s^2+4.2s+9)(s+64.9)(s+26.5)(s+2.4)} \tag{51}$$

Other Approaches to Synthesis

As pointed out at the start, one should always endeavor to transform the characteristic equation by cancellation or some other means in such a way as to make the unknown parameters figure only in the co-ordinates of the working point $M(a_1, a_0)$. In the preceding examples it was shown how this can be achieved in some practical cases. It is not to be understood, however, that this is the only way possible. First, if the unknown parameters must be present in the coefficient a_3, it is possible to construct the curves Γ_t for different values of the coefficient a_3 (usually 3 or 4) and solve the synthesis problem by means of interpolation with respect to a_3. The same applies for the case when one parameter is present in several coefficients.

Other possibilities also exist: some of them will be briefly explained. One

possibility is the separate synthesis of small and large poles of the closed-loop transfer function. It is well known that the poles whose real part is large compared with the real part of control poles have negligible effect on the unit step function response of the system. This fact suggests the possibility of selecting separately large and small poles during the synthesis. The synthesis could then be started with small poles and then completed with the large ones, or conversely. To explain this more specifically take the characteristic equation

$$a_n s^n + a_{n-1} s^{n-1} + \ldots + a_2 s^2 + a_1 s + a_0 = 0 \quad (52)$$

For small roots, the values s^4, s^5, $\ldots$ s^n are small compared with the lower powers, and they can be neglected. The small roots of equation 52 are therefore approximately determined by the equation

$$a_3 s^3 + a_2 s^2 + a_1 s + a_0 \approx 0 \quad (53)$$

Similarly, large roots of equation 52 can be approximately determined from the equation

$$a_n s^n + a_{n-1} s^{n-1} + a_{n-2} s^{n-2} + a_{n-3} s^{n-3} = 0 \quad (54)$$

Even better approximation could be achieved if equations of the fourth degree were used instead of equations 53 and 54. The calculation did not go further than the third degree because the construction of the curves Γ_ξ can be avoided by using the curves in Figs. 5 and 6, which are already available. In some cases it may be sufficient to adopt the second degree for one or both of equations 53 and 54.

Take for example the characteristic equation of the fifth degree

$$a_5 s^5 + a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0 = 0 \quad (55)$$

If the synthesis is started with small roots, the first two terms may be neglected, and instead of equation 55 we obtain equation 53. Using the curves in Figs. 5 and 6, one can then perform the synthesis as though the system in question had a characteristic equation of the third degree. Thereby the coefficients a_0, a_1, a_2, and a_3 would be determined as well as all parameters which take part in them. If the values of these coefficients are then introduced in equation 55, the only unknown values will be a_4 and a_5. These two coefficients will become the coordinates of the working point M if equation 55 is transformed by the substitution

$$s = \frac{1}{p} \quad (56)$$

into the equation

$$a_5 p^5 + a_4 p^4 + a_3 p^3 + a_2 p^2 + a_1 p + a_0 = 0 \quad (57)$$

One should then construct the corresponding curves Γ_ξ and solve finally the design problem by selecting the working point $M(a_4, a_5)$ taking into account that equation 56 yields $\omega_{nt} = 1/\omega_n$ and $\zeta_t = \zeta$, where the subscript t denotes the magnitude belonging to variable p. If the given characteristic equation were of a higher degree it would be possible, by repeating the procedure applied for a_4 and a_5 to determine the pairs a_6 and a_7, a_8 and a_9, etc.

Similarly, if the synthesis of a system with the characteristic equation 55 were to be started with large roots, one should first apply equation 54 and determine the coefficients a_5, a_4, a_3 and a_2 in such a way that the negative real parts of all roots are large compared with the real part of control poles. It would also be desirable if these large roots could all be real. It might then be possible to obtain the final system with only one complex pair of poles. Once the numerical values of the determined coefficients are substituted in equation 55, there remain only a_1 and a_0 as unknown values. The subsequent procedure consists in selecting the position of the working point $M(a_1, a_0)$ in accordance with the specified requirements. The large roots would thereby be slightly altered but any influence they may have is negligible. However, the procedure would be most appropriate if the synthesis were independently started with both large and small roots.

Another basically different way of solving the synthesis problem of feedback control system is also possible. Instead of using the open-loop transfer function as a starting point in adjusting the poles of the closed-loop transfer function, it is possible to start from the latter and adjust the poles of the open-loop transfer function.

If the closed-loop transfer function of a simple system with only one feedback loop is given in the form 1, the poles of the open-loop transfer function are equal to the roots of the algebraic equation

$$f(s) - p(s) = 0 \quad (58)$$

With the use of this equation and the associated curves Γ_0, Γ_1, and Γ_ξ the synthesis problem of a feedback control system can in some cases be handled by selecting a pair of control poles of the closed-loop transfer function, according to the specified system characteristics, and then seeking a rational function by which this basic part is to be multipled in order to yield the desired (or specified) open-loop transfer function. This will be demonstrated in the following example.

The transfer function of the fixed part of the system is

$$G_2(s) = \frac{K_2}{s(s+a)} \quad (59)$$

The desired characteristics of the system are determined by the specifications concerning the bandwidth B, the damping coefficient of the control poles ζ, the corresponding overshoot in per cent, and the velocity constant K_v. Since the solution to be obtained is of a general nature, the numerical values are not given.

First, the control poles of the closed-loop transfer function will be chosen. Their damping coefficient being specified, it only remains to assume an undamped natural frequency ω_n. Since the bandwidth B is also specified, one determines ω_n from the well-known relations between B and ω_n, bearing in mind that a zero of the closed-loop transfer function will also appear.

Now, when $\lambda_1 = 2\omega_n\zeta$ and $\lambda_0 = \omega_n^2$, are substituted, the closed-loop transfer function takes the form

$$\frac{C}{R}(s) = \frac{K}{s^2 + \lambda_1 s + \lambda_0} \times \frac{s+\alpha}{s+\beta} \quad (60)$$

where λ_1 and λ_0 are known, while α and β are to be determined according to specifications. To simplify the compensation transfer function it would be suitable to choose α and β in such a way as to make the two poles of the open-loop transfer function corresponding to the closed-loop transfer function 60 equal to the poles of the transfer function $G_2(s)$. The poles of the open-loop transfer function are determined by the equation

$$s^3 + (\lambda_1 + \beta)s^2 + (\lambda_0 + \beta\lambda_1 - K)s + \beta\lambda_0 - K\alpha = 0 \quad (61)$$

which, by introducing $s = (\lambda_1 + \beta)p$, is transformed into

$$p^3 + p^2 + \frac{\lambda_0 + \beta\lambda_1 - K}{(\lambda_1 + \beta)^2} p + \frac{\beta\lambda_0 - K\alpha}{(\lambda_1 + \beta)^3} = 0 \quad (62)$$

The curve $\Gamma_{1\xi}$ corresponding to this equation is constructed in Fig. 5 or Fig. 6. Now, select the point M in such a way that the two poles of the open-loop transfer function are equal to the poles of $G_2(s)$. This implies that one pole must be zero, another $-a$, and the third real and negative. For one of the roots of equation 62 to be zero, the point M, according to Fig. 5 or Fig. 6, must lie on the ξ axis, resulting in the condition

$$\beta\lambda_0 - K\alpha = 0 \quad (63)$$

The second pole of the open-loop trans-

fer function will be $-a$ if the point M lies on the tangent whose slope is $\omega_{n1}=a(\lambda_1+\beta)$. On account of this and the preceding condition, the point M must coincide with the intersection of the tangent and the ξ axis. The abscissa of this point, determined from equation 38 of Part I, must be equal to the coefficient of p in equation 62, yielding the condition

$$\frac{\lambda_0+\beta\lambda_1-K}{(\lambda_1+\beta)^2}=\frac{a}{\lambda_1+\beta}-\frac{a^2}{(\lambda_1+\beta)^2} \tag{64}$$

which, when multiplied by $(\lambda_1+\beta)^2$ and rearranged, becomes

$$K-(\lambda_1-a)\beta=\lambda_0-a\lambda_1+a^2 \tag{65}$$

With the use of relation 23 the specification concerning the velocity constant yields the condition

$$\frac{1}{\alpha}-\frac{1}{\beta}=\frac{\lambda_1}{\lambda_0}-\frac{1}{K_v} \tag{66}$$

Equations 63, 65, and 66 form a system of three equations with three unknowns. When solved, this system gives the values of α, β, and K, for which the open-loop transfer function has two poles equal to the poles of $G_R(s)$. At the same time, all performance specifications of the system will also be satisfied. The solutions are

$$\alpha=\frac{a}{1+\dfrac{2\omega_n\zeta-\omega_n^3/K_v}{a-2\omega_n\zeta}+\left(2\omega_n\zeta-\dfrac{\omega_n^3}{K_v}\right)\dfrac{a}{\omega_n^2}} \tag{67}$$

$$\beta=\frac{a}{1+\dfrac{2\omega_n\zeta-\omega_n^3/K_v}{a-2\omega_n\zeta}} \tag{68}$$

$$K=\omega_n^2\frac{\beta}{\alpha} \tag{69}$$

With these values, one pole of the open-loop transfer function will be zero, another $-a$, and the third is obtained from 61 by first eliminating the other two. Thus $p_3=2\omega_n\zeta+\beta-a$ is obtained and the open-loop transfer function can be written in the form

$$G(s)=\frac{K(s+\alpha)}{s(s+a)(s+2\omega_n\zeta+\beta-a)} \tag{70}$$

By comparing equations 70 and 59, while leaving out of consideration the constant term to be realized by amplifier, one obtains the compensation transfer function in the following form:

$$G_1(s)=\frac{s+\alpha}{s+2\omega_n\zeta+\beta-a} \tag{71}$$

It is important to bear in mind that relations 67 and 68 may yield negative values for α and β, which is obviously unacceptable. Hence it is necessary that $\alpha>0$ and $\beta>0$, which results in the following conditions for the use of equations 67, 68, and 69:

1. If $a-2\omega_n\zeta>0$ or $a-2\omega_n\zeta<-\dfrac{\omega_n^2}{a}$, it is necessary that

$$2\omega_n\zeta-\frac{\omega_n^3}{K_v}>-\frac{a-2\omega_n\zeta}{1+(a-2w_n\zeta)\dfrac{a}{\omega_n^2}} \tag{72}$$

2. If $0>a-2\omega_n\zeta>-\dfrac{\omega_n^2}{a}$, the following condition is pertinent:

$$2\omega_n\zeta-\frac{\omega_n^3}{K_v}<-(a-2\omega_n\zeta) \tag{73}$$

Now, take the following numerical example: The fixed part of the system is described by the transfer function of the form 59 where $a=10$. The overshoot should be less than 8%, the bandwidth β less than 60 radians/sec and the velocity constant greater than 50/sec. Owing to the presence of one zero and one pole whose values are still unknown, it is better to be conservative and to assume $\zeta=0.707$ and $\omega_n=30$. Since condition 73 is satisfied, the equations for α, β, and K, may be applied which yield $\beta=40.5$, $\alpha=19.28$, and $K=1{,}890$. According to equations 60, 70, and 71 the corresponding transfer function will be

$$\frac{C}{R}(s)=\frac{1{,}890(s+19.28)}{(s^2+42.4s+900)(s+4.05)};$$

$$G(s)+\frac{1{,}890(s+19.28)}{s(s+10)(s+72.9)};\ G_1(s)=\frac{s+19.28}{s+72.9} \tag{74}$$

The same equations may be applied in certain cases where the transfer function of the fixed part possesses a double zero at infinity. Such transfer functions, for instance, are

$$G_R(s)=\frac{K_3(s+b)}{s(s^2+\lambda_1 s+\lambda_0)};\ G_R(s)=\frac{K_3(s+b)}{s(s+c_1)(s+c_2)} \tag{75}$$

The corresponding compensation transfer functions are

$$G_1(s)=\frac{(s^2+\lambda_1 s+\lambda_0)(s+a)}{(s+a)(s+2\omega_n\zeta+\beta-a)(s+b)};$$

$$G_1(s)=\frac{(s+c_1)(s+c_2)(s+\alpha)}{(s+a)(s+2\omega_n\zeta+\beta-a)(s+b)} \tag{76}$$

Here the choice of the parameter a is arbitrary, though to some extent it may be governed by conditions 72 and 73 and/or by better realizability of the transfer function 76.

Though this type of reverse feedback control system synthesis is very effective in the case of a transfer function of the form 59, it is not of general validity. The basic reason is that the poles of the transfer function of the fixed part are more rigorously specified and the adjustment is therefore considerably more difficult. This reverse procedure should therefore be understood as an illustration of the flexibility of the method rather than a general instruction for its application.

The possibilities of the method are actually far greater than is shown here. If, for instance, equation 5 of Part I is taken as the locus of points corresponding to the roots with constant undamped natural frequency ω_n, the variables ξ and η will be linearly dependent on the function $\phi_2(\xi)$. Since these functions are given in Table I of Part I, the calculation of the curves corresponding to the selected ω_n involves no difficulties. These curves can also be used in the analysis and synthesis of control systems, both separately and conjointly with the curves Γ_0, Γ_1, and Γ_ξ.

One can also easily construct the curves corresponding to constant settling time ($\omega_n\zeta=$ constant). These curves may be of value in certain problems.

References

1. Graphical Analysis and Synthesis of Feedback Control Systems, Part I, Theory and Analysis, Dušan Mitrović. *AIEE Transactions*, pp. 476–87 of this issue.

2. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

Graphical Analysis and Synthesis of Feedback Control Systems

III—Sampled-Data Feedback Control Systems

DUŠAN MITROVIĆ
NONMEMBER AIEE

EXCEPT FOR some special methods, two main approaches are possible in the analysis of sampled-data feedback control systems: 1. Analysis in p plane, p being the complex variable of Laplace transformation;[1] and 2. Analysis in z-plane, z being the complex variable of Z-transformation.[2]

The poles of the corresponding closed-loop system function are significant in both approaches. In the case 1 they are determined by an algebraic equation in e^{pT}, T being the sampling period; and in the case 2 by an algebraic equation in z.

The method proposed in this paper is based on the same principles as presented in references 3 through 6, i.e., it departs from the system algebraic equation, the two last coefficients of which are considered as variables ξ and η. Thus the problem is transferred to the real plane $0\xi\eta$, which greatly simplifies the analysis. Taking into account that the degree and the coefficients of the two aforementioned system algebraic equations are equal, it follows that the same $0\xi\eta$ plane and the same curves apply to both p-plane and z-plane analysis.

Application in the analysis and synthesis of sampled-data feedback control systems will be illustrated by examples. Of especial interest will be the stability analysis; the design procedure when the settling time or the relative damping ratio are specified; and the system compensation.

Since the poles of the closed-loop z transfer function are obtained during the study of a given problem, it is possible, by applying the inverse z transformation, to obtain the correct values of the output time function at the sampling instants. The result is obtained without any approximation, implying that the possible inaccuracies may originate only from the graphical errors.

Given the algebraic equation in e^{pT}, then

$$a_n e^{npT} + a_{n-1} e^{(n-1)pT} + \ldots + a_2 e^{2pT} + a_1 e^{pT} + a_0 = 0 \quad (1)$$

where all coefficients a_k are real and T is positive. By substituting

$$z = e^{pT} \quad (2)$$

equation 1 becomes

$$a_n z^n + a_{n-1} z^{n-1} + \ldots + a_2 z + a_1 z + a_0 = 0 \quad (3)$$

If a sampled-data feedback control system is to be analyzed in p-plane, the system equation will be of the form 1. In the z-plane analysis the form 3 appears. For a given system the degree and the coefficients are equal in both cases.

Considering the coefficients a_1 and a_0 as variables ξ and η, and introducing in equation 3

$$z = -\rho_z \zeta_z + j\rho_z \sqrt{1-\zeta_z^2} \quad (4)$$

the following relations are obtained

$$\xi = a_2 \phi_2(\zeta_z) \rho_z + a_3 \phi_3(\zeta_z) \rho_z^2 + \ldots$$
$$\eta = -\rho_z^2 [a_2 \phi_2(\zeta_z) + a_3 \phi_3(\zeta_z) \rho_z + \ldots] \quad (5)$$

As shown in reference 5, equations 5 may represent the loci of points in the $0\xi\eta$-plane corresponding to the roots of equation 3 with ζ_z = constant or ρ_z = constant.

In a similar way, by introducing in equation 1 the substitution

$$p = -\rho\zeta + j\rho\sqrt{1-\zeta^2} \quad (6)$$

it is possible to express the variables ξ and η as functions of the relative damping coefficient ζ and the undamped natural frequency ρ. Indeed, the same result is more easily obtained by help of the relations

$$\rho_z = e^{-\rho\zeta T} \quad (7)$$

$$\zeta_z = -\cos \rho T \sqrt{1-\zeta^2} \quad (8)$$

which follow directly from equations 2, 4 and 6. On account of equations 7 and 8, the equations 5 become

$$\xi = a_2 \phi_2(-\cos \rho T \sqrt{1-\zeta^2}) e^{-\rho\zeta T} + a_3 \phi_3(-\cos \rho T \sqrt{1-\zeta^2}) e^{-2\rho\zeta T} + \ldots$$
$$\eta = -e^{-2\rho\zeta T} [a_2 \phi_2(-\cos \rho T \sqrt{1-\zeta^2}) + a_3 \phi_3(-\cos \rho T \sqrt{1-\zeta^2}) e^{-\rho\zeta T} + \ldots] \quad (9)$$

These equations may represent the loci of points in the $0\xi\eta$-plane corresponding to the roots with constant damping coefficient, with constant undamped natural frequency or with constant settling time, depending on which variable among ζ, ρ, and $\rho\zeta$ is considered as constant. Calculation of the points of these curves represents no particular difficulty since the table of function e^{-x} and the following table of functions shown as Table I may be used.

Since the coefficients a_1 and a_0 are the same in equations 1 and 3, the variables ξ and η in equations 5 and 9 have the same meaning. Consequently, if the same condition is introduced in equations 5 and 9, the resulting equations must represent the same curves. The choice depends only on the problem to be solved. In other words, the same $0\xi\eta$-plane corresponds to both p-plane and z-plane.

Stability Analysis

Consider the single-loop unity feedback closed-loop system shown in Fig. 1. For other types of sampled-data feedback control systems the analysis is similar and the procedure to be explained is applicable without any essential modification.

Stability of the system of Fig. 1 can be determined by investigating the location of the zeros of $1+y_{(p)}^*$, where the starred transfer function $y_{(p)}^*$ is a periodic function with a period $j\omega_s = j2\pi/T$. Consequently, if all zeros of $1+y_{(p)}^*$, corresponding to the primary period are located in the left half-plane, the stability of the system will be ensured. Since $1+y_{(p)}^*$ is a function with real coefficients, it is necessary to investigate only the region shown shaded in Fig. 2.

Zeros of $1+y_{(p)}^*$ are determined by an equation of the form 1. In the $0\xi\eta$-plane analysis the system will be stable if the working point $M(a_1, a_0)$ is located inside the stable region. The latter is bounded by the curves corresponding to the straight lines limiting the shaded region in Fig. 2. It is, indeed, easier to obtain the equation of these curves by departing from the analysis in z-plane. In this case the system will be stable if all zeros of $1+y_{(z)}^*$ are inside the unit circle, since the transformation $z = e^{pT}$ maps the left half of the p-plane into the interior of the unit circle in the z-plane.

Determination of zeros of $1+y_{(p)}^*$ is

Paper 58-862, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 23-27, 1958. Manuscript submitted November 1, 1957; made available for printing April 22, 1958.

DUŠAN MITROVIĆ is with the University of Belgrade and the Boris Kidritch Institute, Belgrade, Yugoslavia.

Table I. Numerical Values of Functions $\phi_k(-\cos x)$

x	$\zeta_x=-\cos x$	$\phi_1(\zeta_x)$	$\phi_2(\zeta_x)$	$\phi_3(\zeta_x)$	$\phi_4(\zeta_x)$	$\phi_5(\zeta_x)$	$\phi_6(\zeta_x)$	$\phi_7(\zeta_x)$	$\phi_8(\zeta_x)$	$\phi_9(\zeta_x)$	$\phi_{10}(\zeta_x)$	$\zeta_x=-\cos x$	x
1.570 ($\pi/2$)	0.00	−1	0.0	+1.00	0.000	−1.000	0.000	+1.000	0.000	−1.000	0.000	0.00	1.570 ($\pi/2$)
1.620	0.05	−1	0.1	+0.99	−0.199	−0.970	+0.296	+0.940	−0.390	−0.901	+ 0.480	−0.05	1.520
1.670	0.10	−1	0.2	+0.96	−0.392	−0.882	+0.568	+0.768	−0.722	−0.623	+ 0.847	−0.10	1.470
1.720	0.15	−1	0.3	+0.91	−0.573	−0.738	+0.794	+0.500	−0.944	−0.216	+ 1.009	−0.15	1.420
1.780	0.20	−1	0.4	+0.84	−0.736	−0.546	+0.954	+0.164	−1.020	+0.244	+ 0.922	−0.20	1.368
1.828	0.25	−1	0.5	+0.75	−0.875	−0.313	+1.031	−0.203	−0.930	+0.668	+ 0.586	−0.25	1.318
1.876	0.30	−1	0.6	+0.64	−0.984	−0.050	+1.014	−0.559	−0.679	+0.966	+ 0.099	−0.30	1.266
1.928	0.35	−1	0.7	+0.51	−1.057	+0.230	+0.896	−0.857	−0.296	+1.064	− 0.449	−0.35	1.212
1.982	0.40	−1	0.8	+0.36	−1.088	+0.510	+0.680	−1.054	+0.164	+0.923	− 0.902	−0.40	1.140
2.037	0.45	−1	0.9	+0.19	−1.071	+0.774	+0.374	−1.111	+0.625	+0.548	− 1.119	−0.45	1.103
2.094	0.50	−1	1.0	+0.00	−1.000	+1.000	0.000	−1.000	+1.000	0.000	− 1.000	−0.50	1.047
2.154	0.55	−1	1.1	−0.21	−0.869	+1.166	−0.413	−0.711	+1.196	−0.604	− 0.531	−0.55	0.987
2.214	0.60	−1	1.2	−0.44	−0.672	+1.246	−0.824	−0.258	+1.133	−1.102	+ 0.189	−0.60	0.926
2.278	0.65	−1	1.3	−0.69	−0.403	+1.214	−1.175	+0.314	+0.767	−1.311	+ 0.937	−0.65	0.862
2.346	0.70	−1	1.4	−0.96	−0.056	+1.038	−1.398	+0.918	+0.112	−1.075	+ 1.393	−0.70	0.785
2.418	0.75	−1	1.5	−1.25	+0.375	+0.687	−1.406	+1.422	−0.727	−0.332	+ 1.225	−0.75	0.722
2.498	0.80	−1	1.6	−1.56	+0.806	+0.126	−1.098	+1.631	−1.511	+0.787	+ 0.252	−0.80	0.644
2.585	0.85	−1	1.7	−1.89	+1.513	−0.682	−0.353	+1.283	−1.828	+1.824	− 1.273	−0.85	0.555
2.692	0.90	−1	1.8	−2.24	+2.232	−1.775	+0.968	+0.036	−1.032	+1.822	− 2.247	−0.90	0.450
2.822	0.95	−1	1.9	−2.61	+3.059	−3.202	+3.025	−2.545	+1.811	−0.896	+ 0.109	−0.95	0.318
3.140 (π)	1.00	−1	2.0	−3.00	+4.000	−5.000	+6.000	−7.000	+8.000	−9.000	+10.000	−1.00	0.000

Instructions for use: The values of functions $\phi_k(-\cos x)$ are given only for x between $\pi/2$ and π. For x between 0 and $\pi/2$ the same values are valid, but the values of functions ϕ_3, ϕ_5, ϕ_7, ϕ_9 and ϕ_{10} change their sign.

effected by solving the algebraic equation 3 and the stability is determined by assuring oneself that all roots of this equation are inside the unit circle. The same statement, expressed in terms of analysis in the $0\xi\eta$-plane, has the following meaning.

The stable region in the $0\xi\eta$-plane is bounded the following:

1. The locus of points corresponding to the roots with constant $\rho_x=1$, which is the limit curve for complex roots.

2. The locus of points corresponding to positive unit roots, which is the limit curve for positive roots.

3. The locus of points corresponding to negative unit roots, which is the limit curve for negative roots.

The equations of locus 1 are easily obtained by putting in equation 5 $\rho_z=1$:

$$\left.\begin{aligned}\xi&=a_0\phi_2(\zeta_x)+a_2\phi_3(\zeta_x)+\dots\\ \eta&=-a_1\phi_1(\zeta_x)-a_3\phi_4(\zeta_x)-\dots\end{aligned}\right\}(\Gamma^1) \qquad (10)$$

where ζ_x is the variable parameter, and the values of function $\phi_k(\zeta_x)$ are given in Table I. Consequently, for the points of curve (Γ^1) to be calculated, the cumulative multiplication is all that is necessary. (The curves and the straight lines are designated as follows: (Γ^{ρ_x}) for $\rho_x=$ constant and (Γ_ζ^1)* for $\zeta=$constant.)

If equations 10 were to be deduced by departing from the analysis in p-plane, the expression $-\cos\rho T$ instead of ζ_x would be involved. Indeed, for $\zeta=0$ equation 8 gives $\zeta_x=-\cos\rho T$, showing that the same curves are obtained in both approaches.

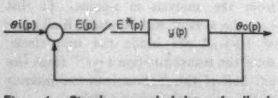


Fig. 1. Simple sampled-data feedback control system

The locus of points corresponding to positive unit roots is the tangent to the curve (Γ_1) having the slope -1. Its equation may simply be written by setting $z=-1$ in equation 3 and by taking coefficients a_1 and a_0 as variable co-ordinates ξ and η.[5] Thus is obtained

$$\eta=\xi-a_2+a_3-a_4+\dots-(-1)^n a_n \quad (L^1) \qquad (11)$$

Similarly, the locus of item 3 is the straight line (L^{-1}) defined by the equation

$$\eta=-\xi-a_2-a_3-a_4-\dots-a_n \quad (L^{-1}) \qquad (12)$$

Now, the condition of absolute stability can be expressed in the following manner:

Sampled-data feedback control system the characteristics of which are determined by equation 1 or equation 3 will be stable if the working point $M(a_1, a_0)$ is located inside the stable region bounded by the curve (equation 10) and the straight lines (equations 11 and 12).

The practical application of this criterion is very simple. This will be shown by examples.

Consider the system in Fig. 1, the conventional linear part of which is characterized by the transfer function

$$y_{(p)}=\frac{K}{p(p+p_1)} \qquad (13)$$

The gain K required to place the system on the dividing lines between stability and instability is to be determined.

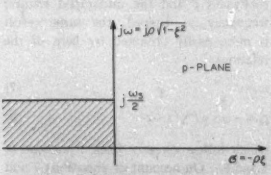


Fig. 2. Region to be investigated

The z transfer function corresponding to $y_{(p)}$ is easily obtained:

$$y_{(z)}{}^*=\frac{K}{p_1}\,\frac{(1-\epsilon^{-p_1T})z}{(z-1)(z-\epsilon^{-p_1T})} \qquad (14)$$

Zeros of $1+y_{(z)}{}^*$ are determined by equation

$$z^2+\left[\frac{K}{p_1}(1-\epsilon^{-p_1T})-(1+\epsilon^{-p_1T})\right]z+\epsilon^{-p_1T}=0 \qquad (15)$$

On account of equations 10, 11, 12, and 15 the characteristic curves are

$$\left.\begin{aligned}\xi&=2\zeta_x\\ \eta&=1\end{aligned}\right\}(\Gamma^1) \qquad (16)$$

$$\eta=\xi-1 \quad (L^1) \qquad (17)$$

$$\eta=-\xi-1 \quad (L^{-1}) \qquad (18)$$

Since ζ_x varies only between -1 and 1, the equation 16 represents the straight line between A and B in Fig. 3. Triangle ABC represents the stable region. The system will be stable if the working point $M[(K/p_1)(1-\epsilon^{-p_1T})-(1+\epsilon^{-p_1T}), \epsilon^{-p_1T}]$ is located inside this triangle. If K is allowed to vary, the stable region remains unchanged, but the point M moves along the straight line

$$\eta=\epsilon^{-p_1T} \qquad (19)$$

which may be labeled by the values of K. Now, it is easy to determine graphically

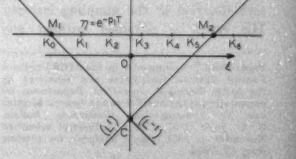


Fig. 3. Stable region for a second-degree sampled-data feedback control system

Table II. Calculation of Curves for System Defined by Equation 26

K	Equations	f_2	ϕ_1	ϕ_2	ξ	η
0	$\xi=-1.154\phi_1+\phi_2$	1	2	−3	−5.308	−3.154
	$\eta=-1.154-\phi_1$	0.5	1	0	−1.154	−2.154
	$\eta=\xi+2.154$	0	0	1	1.0	−1.154
	$\eta=-\xi+0.154$	−0.5	−1	0	1.154	−0.154
	M (0.150; −0.00248)	−1	−2	−3	−0.692	0.846
4	$\xi=0.342\phi_1+\phi_2$	1	2	−3	−2.316	−1.658
	$\eta=0.342-\phi_1$	0.5	1	0	0.342	−0.658
	$\eta=\xi+0.658$	0	0	1	1.0	0.342
	$\eta=-\xi-1.342$	−0.5	−1	0	−0.342	1.342
	M (0.3584; −0.00248)	−1	−2	−3	−3.684	2.342
8	$\xi=1.838\phi_1+\phi_2$	1	2	−3	0.676	−0.162
	$\eta=1.838-\phi_1$	0.5	1	0	1.838	0.838
	$\eta=\xi-0.838$	0	0	1	1.0	1.838
	$\eta=-\xi-2.838$	−0.5	−1	0	−1.838	2.838
	M (0.5608; −0.00248)	−1	−2	−3	−6.676	3.838

those values of K at which the system becomes unstable. In this simple case the analytical solution is also possible. It consists in determining the points of intersections M_1 and M_2 by the help of equations 17, 18, and 19. Thus are obtained the relations

$$\frac{K}{p^1}(1-\epsilon^{-p_1T})-(1+\epsilon^{-p_1T})=-(1+\epsilon^{-p_1T}) \quad (20)$$

$$\frac{K}{p_1}(1-\epsilon^{-p_1T})-(1+\epsilon^{-p_1T})=1+\epsilon^{-p_1T} \quad (21)$$

When solved for K, these equations determine the following stable interval

$$0<K<2p_1\frac{1+\epsilon^{-p_1T}}{1-\epsilon^{-p_1T}} \quad (22)$$

The next example will again be the system of Fig. 1 but having the transfer function

$$y(p)=\frac{K}{p(p+1)(p+2)} \quad (23)$$

and the corresponding z transfer function

$$y(z)^*=K\times\frac{\left(\frac{\epsilon^{-2T}}{2}-\epsilon^{-T}+\frac{1}{2}\right)z^2+\left(\frac{\epsilon^{-2T}}{2}-\epsilon^{-2T}+\frac{1}{2}\epsilon^{-T}\right)z}{(z-1)(z-\epsilon^{-T})(z-\epsilon^{-2T})} \quad (24)$$

Zeros of $1+y(z)^*$ are determined by equation

$$z^3+\left[K\left(\frac{\epsilon^{-2T}}{2}-\epsilon^{-T}+\frac{1}{2}\right)-(\epsilon^{-2T}+\epsilon^{-T}+1)\right]z^2+\left[K\left(\frac{\epsilon^{-2T}}{2}-\epsilon^{-2T}+\frac{1}{2}\epsilon^{-T}\right)+(\epsilon^{-2T}+\epsilon^{-2T}+\epsilon^{-T})\right]z-\epsilon^{-2T}=0 \quad (25)$$

For a prescribed sampling frequency of π rad/sec, this equation becomes

$$z^3+(0.374K-1.154)z^2+(0.0506K+0.156)z-0.00248=0 \quad (26)$$

The stable region is bounded by the curves

$$\left.\begin{aligned}\xi&=(0.374K-1.154)\phi_2(f_2)+\phi_3(f_2)\\ \eta&=(0.374K-1.154)-\phi_2(f_2)\end{aligned}\right\}(\Gamma^1) \quad (27)$$

$$\eta=\xi-0.374K+2.154 \quad (L^1) \quad (28)$$

$$\eta=-\xi-0.374K+0.154 \quad (L^{-1}) \quad (29)$$

and the working point $M(0.0506K+0.156,\ -0.00248)$.

Equations 27, 28, and 29 represent the systems of curves depending on K. Since the construction of parallel straight lines (equations 28 and 29) presents no difficulty, the problem is reduced to the construction of curves 27. Table I is to be used for this purpose. The construction of the curves (Γ^1) for a few values of K can thus be easily performed. To give an idea of the amount of work involved, the entire calculation is given in Table II.

Curves (Γ_1) and the straight lines (L^1) and (L^{-1}) for three values of K are constructed in Fig. 4. As K varies, the working point M moves along a straight line parallel to the $\xi-$axis, which is labeled with numbers 0, 4, and 8, corresponding to the respective values of K. When $K=0$, the working point M is on the straight line (L^{-1}). For $K=4$ it is located in the stable region shown shaded in Fig. 4, but for $K=8$ it is already outside. Evidently the working point M has left the stable region by crossing the straight line (L^1). By interpolation between the straight lines (L^1) for $K=4$ and $K=8$ the greatest allowable $K=7$ is easily obtained. Indeed, this value can also be determined analytically by using the condition that M must lie on the straight line (L^1). Thus is obtained the equation

$$-0.00248=0.0506K+0.156-0.374K+2.154 \quad (30)$$

which gives $K\approx7.15$. Consequently, the system will be stable for all values of K belonging to the interval $0<K<7.15$.

Settling Time Region

Settling time is the time required for the system to settle to an essentially steady-state value after the application of a step input function.[7] It is determined by the time constant of the envelope of the damped oscillation, the latter being the reciprocal of the real part of control poles $(1/\rho_1^r)$. Consequently, the design of a sampled-data system with a certain maximum settling time requires that all poles of the system function lie in

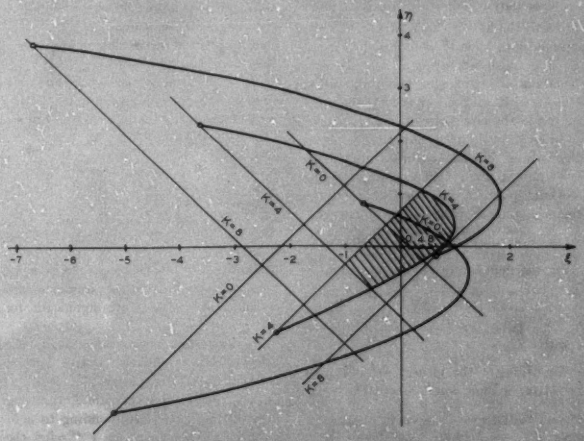


Fig. 4. Stability regions for a third-degree equation

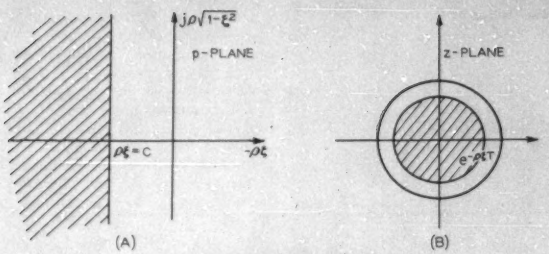


Fig. 5. Loci of constant time constant

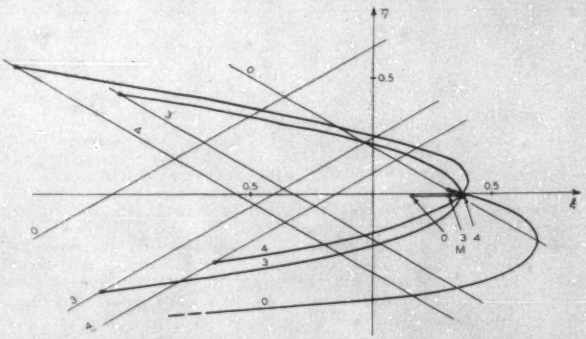


Fig. 6. Determination of the gain K corresponding to a prescribed maximum settling time

shaded region of Fig. 5(A), where $\rho\zeta = c =$ constant is determined by the prescribed maximum settling time. Indeed, the transformation (equation 2) maps a line c units to the left of the $j\rho\sqrt{1-\zeta^2}$ in the p-plane into a circle of radius ϵ^{-cT} in the z plane, as shown in Fig. 5(B). Thus, if a maximum settling time is not allowed to be exceeded, all poles in the z-plane must lie inside the circle of radius ϵ^{-cT}.

In terms of the analysis in the $0\xi\eta$-plane, this means that the settling time will be less than a prescribed maximum if the working point $M(a_1, a_0)$ lies inside the region bounded by the following curves:

1. The locus of points corresponding to the roots with constant $\rho_z = \epsilon^{-cT}$, this being the limit curve for complex roots.
2. The locus of points corresponding to positive roots ϵ^{-cT}.
3. The locus of points corresponding to negative roots $-\epsilon^{-cT}$.

Equations of the locus 1 are obtained by setting in equation 5, $\rho_z = \epsilon^{-cT}$:

$$\left.\begin{aligned}\xi &= a_2\epsilon^{-cT}\phi_2(\zeta_z) + a_4\epsilon^{-3cT}\phi_4(\zeta_z) + \dots \\ \eta &= -\epsilon^{-2cT}[a_2\phi_1(\zeta_z) + a_4\epsilon^{-cT}\phi_3(\zeta_z) + \dots]\end{aligned}\right\} \quad (\Gamma^{\rho_z}) \quad (31)$$

These equations represent the locus of points in the $0\xi\eta$ plane corresponding to a constant settling time. Since ϵ^{-cT} is constant, it is advisable to calculate first the modified coefficients $a_k\epsilon^{-c(k-1)T}$ and then to proceed in the same manner as in the case of equation 10.

By application of the procedure outlined in the preceding section the following equations for the loci 2 and 3 are obtained:

$$\eta = \epsilon^{-cT}\xi - a_2\epsilon^{-2cT} + a_3\epsilon^{-3cT} - a_4\epsilon^{-4cT} + \dots - (-1)^n a_n\epsilon^{-ncT} \quad (32)$$

$$\eta = -\epsilon^{-cT}\xi - a_2\epsilon^{-2cT} - a_3\epsilon^{-3cT} - a_4\epsilon^{-4cT} - \dots - a_n\epsilon^{-ncT} \quad (33)$$

These equations represent the straight lines corresponding to the same constant settling time. They are significant for real s's.

As an example, consider again the sampled-data feedback control system characterized by equation 26. It is required to determine the gain K resulting in a settling time corresponding to $\rho\zeta = c = 0.255$. On account of $\epsilon^{-cT} = 0.6$ the equations 31, 32 and 33 have the form

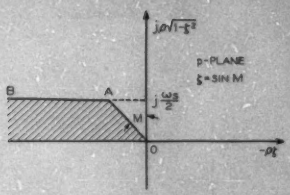


Fig. 7. Region to be investigated when the relative damping is specified

$$\left.\begin{aligned}\xi &= (0.2244K - 0.6924)\phi_2(\zeta_z) + 0.36\phi_4(\zeta_z) \\ \eta &= (0.1346K - 0.4154) - 0.216\phi_3(\zeta_z)\end{aligned}\right\} \quad (\Gamma^{0.6}) \quad (34)$$

$$\eta = 0.6\xi - 0.1346K + 0.6314 \quad (L^{0.6}) \quad (35)$$

$$\eta = -0.6\xi - 0.1346K + 0.1994 \quad (L^{-0.6}) \quad (36)$$

These equations are no more complicated than the corresponding equations 27, 28, and 29. Calculation of points is similar to that presented in Table II, except that it is here performed for $K = 0$, 3, and 4. The curves and the working point $M(0.0506K + 0.156, -0.00248)$ are shown in Fig. 6. The allowable values of K, determined by interpolation, lie in the interval $0.7 < K < 4$.

As is evident from Fig. 6, the working point enters the settling time region by crossing the straight line (equation 36) and leaves it by crossing the curve (equation 34). Consequently, the lower limit for K can also be determined by setting the co-ordinates of M in (equation 36). As for the upper limit for K, the graphical determination is easier, but the result can be checked with the help of equations 34.

Relative Damping Region

If the relative damping is specified, the region shown shaded in Fig. 7 is to be investigated. The relative damping region in the $0\xi\eta$-plane can be found by determining the curves which correspond to the straight lines bounding the aforementioned shaded region. Each limit line will be considered separately.

1. $\overline{OA}$ is the part of the locus of points with constant relative damping $\zeta = d$. The equations of the corresponding locus in the $0\xi\eta$-plane can evidently be found by setting in equation 9 $\zeta = d$. Thus is obtained

$$\left.\begin{aligned}\xi &= a_2\phi_2(-\cos\rho T\sqrt{1-d^2})\epsilon^{-\rho Td} + a_4\phi_4(-\cos\rho T\sqrt{1-d^2})\epsilon^{-2\rho Td} + \dots \\ \eta &= \epsilon^{-2\rho Td}[a_2\phi_1(-\cos\rho T\sqrt{1-d^2}) + a_4\phi_3(-\cos\rho T\sqrt{1-d^2})\epsilon^{-\rho Td} + \dots]\end{aligned}\right\} \quad (\Gamma_\zeta)^* \quad (37)$$

Table III. Calculations Required for Determining Gain K Corresponding to $\zeta>0.5$

k	Equations	$x=\frac{}{2\sqrt{0.75\zeta}}$	$\zeta_a=-\cos$	$\phi_3(\zeta)$	$\phi_3(\zeta_a)$	$\epsilon^{-\rho}$	$\epsilon^{-2\rho}$	$\epsilon^{-3\rho}$	ξ	η
	$\xi=-1.154\phi_3(\zeta_a)\epsilon^{-\rho}+\phi_3(\zeta_a)\epsilon^{-2\rho}$	0.000	−1.0	−2.0	−3.00	1.000	1.000	1.000	−0.692	0.846
	$\eta=-1.154\epsilon^{-2\rho}-\phi_3(\zeta_a)\epsilon^{-3\rho}$	0.926	−0.6	−1.2	−0.44	0.586	0.341	0.201	0.662	−0.182
0	$\eta=0.1628\xi+0.0349$	1.047	−0.5	−1.0	0.00	0.546	0.298	0.162	0.631	−0.182
	$\eta=-\xi+0.154$	1.570	0.0	0.0	1.00	0.403	0.162	0.0654	0.162	−0.187
	M (0.156, −0.00248)	2.094	0.5	1.0	0.00	0.298	0.0888	0.0264	−0.344	−0.129
		3.140	1.0	2.0	−3.00	0.1623	0.0264	0.00428	−0.454	−0.039
	$\xi=-0.78\phi_3(\zeta_a)\epsilon^{-\rho}+\phi_3(\zeta_a)\epsilon^{-2\rho}$	0.000	−1.0	−2.0	−3.00	1.000	1.000	1.000	−1.44	1.22
	$\eta=-0.78\epsilon^{-2\rho}-\phi_3(\zeta_a)\epsilon^{-3\rho}$	0.926	−0.6	−1.2	−0.44	0.586	0.341	0.201	0.399	−0.0245
1	$\eta=0.1628\xi+0.025$	1.047	−0.5	−1.0	0.00	0.546	0.298	0.162	0.426	−0.0703
	$\eta=-\xi-0.220$	1.570	0.0	0.0	1.00	0.403	0.162	0.0654	0.162	−0.1265
	M (0.2066, −0.00248)	2.094	0.5	1.0	0.00	0.298	0.0888	0.0264	−0.2323	−0.0936
		3.140	1.0	2.0	−3.00	0.1623	0.0264	0.00428	−0.3327	−0.0292
	$\xi=-0.406\phi_3(\zeta_a)\epsilon^{-\rho}+\phi_3(\zeta_a)\epsilon^{-2\rho}$	0.000	−1.0	−2.0	−3.00	1.000	1.000	1.000	−2.188	1.594
	$\eta=-0.406\epsilon^{-2\rho}-\phi_3(\zeta_a)\epsilon^{-3\rho}$	0.926	−0.6	−1.2	−0.44	0.586	0.341	0.201	0.136	0.104
	$\eta=0.1628\xi+0.0151$	1.047	−0.5	−1.0	0.00	0.546	0.298	0.162	0.223	0.041
2	$\eta=-\xi-0.594$	1.266	−0.3	−0.6	0.64	0.482	0.232	0.112	0.2665	−0.027
	M (0.2572, −0.00248)	1.570	0.0	0.0	1.00	0.403	0.162	0.0654	0.162	−0.0658
		2.094	0.5	1.0	0.00	0.298	0.0888	0.0264	−0.1211	−0.0635
		3.140	1.0	2.0	−3.00	0.1623	0.0264	0.00428	−0.2112	−0.0193

where ρ is the variable parameter. At first sight the equations 37 appear to be unsuitable for practical application. On the contrary, owing to Table I and the standard table for ϵ^{-x}, the construction of the curve by departing from equation 37 presents no special difficulty, particularly because the part of the curve corresponding to the interval $0\leq\rho T\sqrt{1-\delta^2}\leq\pi$ only is required.

2. $\overline{AB}$ is the part of the straight-line corresponding to constant frequency

$$\rho\sqrt{1-\zeta^2}=\frac{\pi}{T} \tag{38}$$

It follows from equation 38 that

$$\cos\rho T\sqrt{1-\zeta^2}=-1,\ \sin\rho T\sqrt{1-\zeta^2}=0 \tag{39}$$

On account of equations 38, 39, and 6, and of $\zeta=\mathrm{d}$, equation 2 becomes

$$z=-\epsilon^{-\frac{\pi d}{\sqrt{1-d^2}}}=-b \tag{40}$$

For the equation of the locus of points in $0\xi\eta$-plane corresponding to roots $z=-6$ to be determined, the same procedure as in the case of equation 11 is to be applied. Thus is obtained

$$\eta=b\xi-a_2b^2+a_3b^3-\ldots-(-1)^n a_n b^n \quad (L_\xi)^* \tag{41}$$

3. The negative real axis is the locus of points satisfying the condition

$$\rho\sqrt{1-\zeta^2}=0 \tag{42}$$

which leads to $z=1$. Consequently, the corresponding locus in $0\xi\eta$-plane is the straight line defined by equation 12.

The following condition may now be formulated: Relative damping coefficient of a sampled-data feedback control system, characterized by equation 3, will be greater than a prescribed value $\zeta=d$ if the working point $M(a_1, a_0)$ lies inside the region bounded by the curve of equation 37 and the straight lines of equations 41 and 12. This relative damping region must be a part of the stable region.

Especially for $\zeta=1$ the system will be critically damped. In this case equations 37 and 41 become

$$\left.\begin{aligned}\xi&=-2a_2\epsilon^{-\rho T}-3a_3\epsilon^{-2\rho T}-4a_4\epsilon^{-3\rho T}-\ldots\\ \eta&=a_2\epsilon^{-2\rho T}+2a_3\epsilon^{-3\rho T}+3a_4\epsilon^{-4\rho T}+\ldots\end{aligned}\right\}\ (\Gamma_1)^* \tag{43}$$

$$\eta=0 \tag{44}$$

Equations 43 may be simplified by setting $\epsilon^{-\rho T}=\rho_z>0$. Thus, is obtained

$$\left.\begin{aligned}\xi&=-2a_2\rho_z-3a_3\rho_z^2-4a_4\rho_z^3-\ldots\\ \eta&=a_2\rho_z^2+2a_3\rho_z^3+3a_4\rho_z^4+\ldots\end{aligned}\right\}(\Gamma_1)^*\quad \rho_z>0 \tag{45}$$

As an example, determine the gain K for the system characterized by equation 26 to have the relative damping coefficient ζ greater than 0.5. The calculation required for construction is shown in Table III.

The curves and the straight lines for different values of K are constructed in Fig. 8. The working point M lies approximately on the curve $(\Gamma_{0.5})^*$ for $K=2$. Consequently, the relative damping coefficient of the system will be greater than 0.5 if K belongs to the interval $0<K<2$.

Integral Compensation

All necessary relations being already available, an example may now be given. Let be given the sampled-data feedback control system of the type as in Fig. 1 with

$$y(p)=\frac{1}{p(p+1)} \tag{46}$$

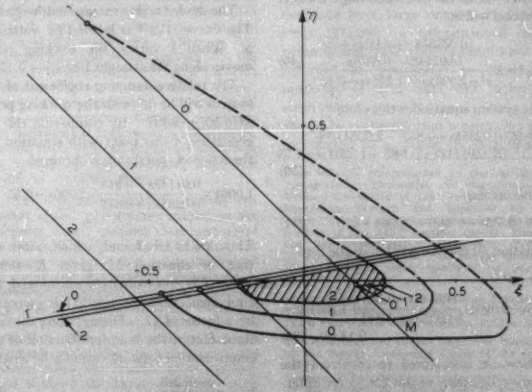


Fig. 8. Determination of K corresponding to $\zeta>0.5$

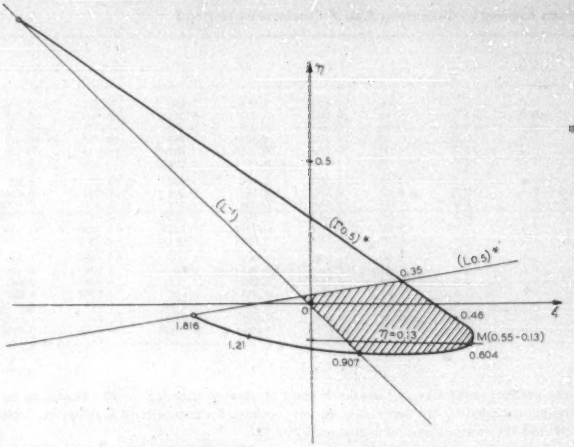


Fig. 9. Relative damping region $\zeta > 0.5$ for $a_2 = -1$

and $\omega_s = \pi$ radians/sec. The equivalent velocity constant for the given $\zeta = 0.5$ is to be increased by a factor of 10. Therefore the integral tandem compensation network is applied, characterized by the transfer function

$$y_{1(p)} = K \frac{p + \lambda\delta}{p + \delta} \tag{47}$$

where δ and $\lambda\delta$ are less than unity and $\lambda > 1$. The transfer function of the conventional linear part of the system is now

$$y_{(p)} = \frac{K(p + \lambda\delta)}{p(p+1)(p+\delta)} \tag{48}$$

With $\delta = 0.02$, the corresponding z transfer function will be

$$y_{(z)^*} = K \frac{(0.02355\lambda + 0.843)z^2 + (0.01142\lambda - 0.843)z}{z^3 - 2.0964z^2 + 1.2264z - 0.13} \tag{49}$$

The system equation is therefore

$$z^3 + [K(0.02355\lambda + 0.843) - 2.0964]z^2 + [K(0.01142\lambda - 0.843) + 1.2264]z - 0.13 = 0 \tag{50}$$

By setting[6]

$$K(0.02355\lambda + 0.843) - 2.0964 = a_2 \tag{51}$$

equation 50 may be written in the form

$$z^3 + a_2 z^2 + \left[(a_2 - 2.0964)\frac{0.01142\lambda - 0.843}{0.02355\lambda + 0.843} + 1.2264\right]z - 0.13 = 0 \tag{52}$$

Now, it is required to construct the relative damping region $\zeta > 0.5$ for different values of a_2. This is done for $a_2 = 0$, -0.5 and -1 but only the curve and the straight lines for $a_2 = -1$ are shown in Fig. 9. Their equations are

$$\left.\begin{aligned} \xi &= -\phi_2(-\cos 2\sqrt{0.75}\times\rho)\epsilon^{-\rho} + \\ &\quad \phi_2(-\cos 2\sqrt{0.75}\rho)\epsilon^{-2\rho} \\ \eta &= -\epsilon^{-\rho} - \phi_2(-\cos 2\sqrt{0.75}\rho)\epsilon^{-2\rho} \end{aligned}\right\} (\Gamma_{0.5})^* \tag{53}$$

$$\eta = 0.1628\xi + 0.0308 \tag{54}$$

$$\eta = -\xi \tag{55}$$

The working point is

$$M\left(1.0964\frac{0.01142\lambda - 0.843}{0.02355\lambda + 0.843} + 1.2264, -0.13\right) \tag{56}$$

The shaded region corresponds to $\zeta > 0.5$. The curve $(\Gamma_{0.5})^*$ is labeled by values of ρ. When λ varies, the working point moves along the straight line $\eta = -0.13$.

The relative damping coefficient of the system will be 0.5 with the working point $M(0.55, -0.13)$. By comparing the coordinates of this point with equation 56, the following condition is obtained

$$1.0964 \times \frac{0.01142\lambda - 0.843}{0.02355\lambda + 0.843} + 1.2264 = 0.55 \tag{57}$$

Thus $\lambda = 12.4$ is found, which, after setting in equation 51, gives $K = 0.966$. Since K is approximately unchanged, the equivalent velocity constant is increased by a factor of 12. Disregarding the constant factor, the transfer function of the compensation network can now be written

$$y_{1(p)} = \frac{p + 0.248}{p + 0.02} \tag{58}$$

If necessary, the poles of the closed-loop z transfer function can easily be determined. The relative damping coefficient ζ of the control poles is already known. The value $\rho = 0.59$ is determined by interpolation on the curve $(\Gamma_{0.5})^*$. Afterwards, equations 7, 8, and 4 give $\rho_z = 0.5543$, $\zeta_z = -0.5225$ and $z_{12} = 0.2894 \pm j0.4825$. The third pole is $z_3 = 0.13/\rho_z^2 = 0.424$. Therefore the closed-loop z transfer function is

$$y_{0(z)^*} = \frac{0.966z(1.135z - 0.705)}{(z^2 - 0.5788z + 0.307)(z - 0.424)} \tag{59}$$

If $y_{0(z)^*}$ is multiplied by the z transform of the input function and the inverse z transformation of this product is evaluated, the correct values of the output time function at the sampling instants can be determined.

It would be possible to obtain the same results without introducing the relation 51. In this case it is necessary to set in equation 50 $K = 1$ and to construct the relative damping region for different values of λ.

References

1. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, W. K. Linvill. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1779–88.

2. The Analysis of Sampled-Data Systems, J. R. Ragazzini, L. A. Zadeh. *Ibid.*, vol. 71, pt. II, Nov. 1952, pp. 225–34.

3. Graphical Conditions for All Roots of an Algebraic Equation to Have Negative Real Parts (in French), D. Mitrović. *Comptes rendus*, Academie des Sciences, Paris, France, vol. 240, 1955, pp. 1177–79.

4. Graphical Conditions for the Arguments of All Roots of an Algebraic Equation to Lie Between $\pi/2 + \mu$ and $3\pi/2 - \mu$ (in French), D. Mitrović. *Ibid.*, vol. 243, 1956, pp. 831–33.

5. Graphical Analysis and Synthesis of Feedback Control Systems, Part I, Theory and Analysis, Dušan Mitrović. *AIEE Transactions*, pp. 476–87 of this issue.

6. Graphical Analysis and Synthesis of Feedback Control Systems, Part II, Syntheses. *Ibid.*, pp. 487–96.

7. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

Discussion

E. I. Jury (University of California, Berkeley, Calif.): This paper extends the methods previously introduced by the author for continuous feedback systems into the sampled-data feedback cases. This extension constitutes a logical procedure since the characteristic equation of the sampled-data case system is algebraic in the variable $z = e^{Ts}$.

It is important to caution readers that this analogy is not always valid. For instance, in examining closely the total response of sampled-data systems, one would encounter entirely different behavior than at the sampling instants.[1] There are

two cases which ought to be investigated before applying the design procedures of this paper: 1. the case of hidden instability, i.e., when the system is stable at the sampling instants but unstable for the total response, and 2. the case of hidden oscillations where the system is stable for all times but the response between sampling instants is oscillatory. In both cases the design for constant damping ratio and of finite settling time introduced by the author are no longer applicable. Indeed, the finite settling time design of the author is applicable only at the sampling instants and in most cases is not applicable to the total response.

The root-locus method in the z-plane can also be advantageously applied to sampled-data systems where the gain could be readily obtained both for stability limit and for constant overshoot or damping ratio.

It might be added that since no hold circuit is used by the author in his examples, the value of the gain K he chooses actually equals the gain of the continuous system multiplied by the pulse width of the sampler h. This stems from the fact that by using the z-transform approximation for the analysis of sampled data, it is assumed the actual finite pulses appearing in the system are mathematically replaced by impulses provided the gain of the continuous system is multiplied by the pulse width. In conclusion, this paper might add more information to the existing information in this field[2] and is useful in stability study and in continuous network compensation of sampled-data control systems.

References

1. Hidden Oscillations in Sampled-Data Control Systems, Eliahu I. Jury, *AIEE Transactions*, vol. 75, pt. II, 1956 (Jan. 1957 section) pp. 391–95.

2. Bibliography of Sampled-Data Control Systems and Z-Transform Applications, H. Freeman, O. Lowenschuss, *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC-4, Mar. 1958.

Dušan Mitrović: I am very much obliged to Mr. Jury for his valuable discussion which draws the attention to certain design problems that must be taken into account. However, the two cases cited in the discussion, as well as the statement concerning the gain, are not problems arising from the author's method, but from the z transformation itself. This is evident if one takes into account that the same problems had been treated before the author's method was proposed (see reference 1 of the discussion).

Analog Computer Simulation of an Aircraft Parallel A-C Generating System

M. RIAZ
MEMBER AIEE

P. E. SMITH, JR.
ASSOCIATE MEMBER AIEE

THE OBJECT of this paper is to present a formulation of the dynamics of complex aircraft electromechanical systems using analytical representations of components suitable for analog computation. The ability of the computer to predict system performance as well as to provide a means for system analysis and synthesis is demonstrated by correlating the results of equipment tests and computer simulation tests in the case of the parallel a-c electrical generating system used in the *B-47* type of aircraft. This comparison is made to insure that all relevant information has been included in the simulation and that the simplifying assumptions adopted in the analysis are reasonably warranted. The approach used to set up the representations is detailed and generalized, with emphasis on the development of an organized procedure for studying complex aircraft electric systems rather than on providing the simplest system representation which may be valid for just one particular case. A system representation using a cascaded sequence of linear gains and time constants with associated feedbacks can give but a crude indication of actual system performance. Similarly, a representation of the dynamics of parallel alternators in terms of equivalent masses and springs, or one using the ordinary stability criteria developed for central station practice, is inadequate.

The computer simulation described in the paper is developed to achieve the following desired objectives:

1. The computer representation must be such as to permit the identification of each component of the system as well as each physical variable characterizing the behavior of the component. Although this approach may add some complexity to the computer arrangement, it has a two-fold advantage: First, it achieves an almost one-to-one correspondence between hardware and computer, thereby permitting easy substitution or changes in components. Second, it provides an organized procedure for setting up the computer and thus facilitates "trouble shooting."

2. The simulation must be carried out so that system performance studies on the computer closely parallel those performed on an actual system. As an example, the computer may have to simulate the following conditions of operation encountered in a two-machine system: excite each isolated alternator at no load; then add rated full load to one alternator; parallel the two generating systems assuming that an automatic paralleling device connects the systems when the relative rotor displacement is about 10 elcetrical degrees; with the system in parallel, suddenly add full load at rated power factor; finally remove the total load. It is with this type of transient operation in mind that analog simulation is to be effected.

3. The simulation must be such that it is possible to examine rapidly the effects of parameter changes on system performance and to assess the advantages and disadvantages of various design modifications.

Set up as specified in the foregoing, the computer simulation establishes a whole new dimension in design procedure and serves as an unexcelled tool for studying various proposed system configurations and for evaluating their relative merits prior to construction and equipment testing. Analog computer studies must, however, also be complemented by other methods before any final assessment of a system can be made. In many instances, digital computation appears to be better adapted to the problems of design and optimization of components which are described by a large number of parameters. Though of unquestioned competence, computer studies should not be regarded as substitutes to hardware, mock-up, or field tests, but rather as steps which precede them. Also, one should be reminded that a computer still serves only as an aid to the system ana-

Paper 58-922, recommended by the AIEE Air Transportation Committee and the Computing Devices Committee of the Communication and Electronics Division, and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting and Air Transportation Conference, Buffalo, N. Y., June 22–27, 1958. Manuscript submitted March 14, 1958; made available for printing April 29, 1958.

M. Riaz is with the Massachusetts Institute of Technology, Cambridge, Mass., and P. E. Smith, Jr., is with Conval Corporation, Cambridge, Mass.

The work described in this paper was conducted by the Energy Conversion Group of the Servomechanisms Laboratory of the Massachusetts Institute of Technology and was supported in part by the Wright Air Development Center under Contract *AF 33(616)3242*.

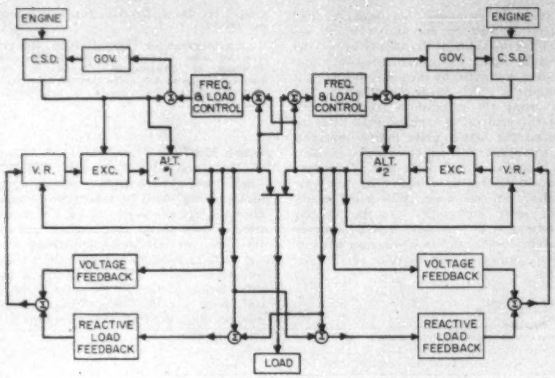


Fig. 1. Configuration of two parallel aircraft a-c generating systems

lyst or designer; it cannot replace his imagination, ingenuity, experience, or intuition.

System Configuration

The specific aircraft system considered in this paper consists of two parallel a-c generating systems arranged as shown in the functional diagram of Fig. 1. Each generating system comprises the following main components:

1. A 50-horsepower Sundstrand constant-speed drive, which basically is a positive-displacement hydraulic transmission. Associated with the drive are a governor and a fine-frequency and load controller which regulate the output speed (and hence the frequency at 400 cps (cycles per second)) and insure proper sharing of real load between the two a-c generators.

2. An alternator (rated 40 kva, 120/208 volts, 0.75 power factor, 6,000 rpm) whose excitation is furnished by a d-c exciter forming an integral part of the machine. The alternator with its exciter is driven through an over-running clutch located at the output of the Sundstrand drive to prevent reversal of power flow when the alternator is tending to act as a motor.

3. A voltage regulator employing a magnetic amplifier whose output is connected to the exciter field while the input receives a d-c signal proportional to the alternator output voltage. The control circuit of the amplifier also has an added signal that is a measure of the difference between the reactive components of currents in each alternator. The circuit is such as to equalize the reactive load between the two parallel machines.

Fig. 1 does not indicate speed, fine frequency, and voltage references as they are considered integral parts of the various controllers.

Block Diagram Representation

The representation of such a complex system by means of block diagrams offers a convenient method for the study of the system dynamic performance. The analytic procedure is to subdivide the complete system into a number of subsystems. Each subsystem normally corresponds to a single physical component. The equations governing the dynamic behavior of each subsystem are established, and a block diagram representation is derived. Finally, the component blocks are interconnected in such a manner that the necessary terminal constraints are satisfied. The block diagrams presented in this paper have the following features:

1. They preserve the identity of all significant physical variables as voltages, currents, torques, and speeds.

2. They include, if only approximately, all nonlinear effects which influence the system behavior. In particular, saturation, variable-speed operation, backlash, two-mode operation (for instance, a clutch), and the like, are all factors represented in the block diagram.

3. Since a block diagram reveals a considerable similarity to an actual computer setup diagram, it is desirable to construct it solely in terms of integration. Differentiation is a noise-amplifying process which should be avoided in an electronic analog computer.

Because several components of the generating system operate at a high power level, it becomes important to consider the loading or coupling effects in the representation. This is accomplished by characterizing such physical components by means of a certain number of terminal ports through which power is either delivered or received. To each terminal port are associated two signals that appear as an input and an output in the component block diagram. The product of this pair of oppositely directed signals has the dimension of power. Consider, for example, the representation of two subsystems A and B, in which subsystem A supplies power to subsystem B acting as a load. The block diagram of A will then furnish an output signal, a voltage, say, which appears as input to the block diagram of B. The block diagram of B will in turn provide, as one of its outputs, the associated power signal, a current now, which becomes an input to subsystem A, thereby accounting for the loading action of B on A. The particular choice of direction of flow for the associated pairs of power signals is dictated by the requirement that the block diagrams must be expressed only in terms of integrators. In addition to these power signals appearing in pairs, there are other single signals which may come in or out of a block diagram and which are usually employed to furnish corrective feedback. These signals normally correspond to low power circuitry in which the matching of components is such that loading effects are negligible.

With these considerations in mind, a general representation of the parallel generating system of Fig. 1 can be obtained by means of block diagrams interconnected in the manner shown in Fig. 2. The representations of the component parts of the two generating systems are identical. The block diagram representation of the interconnections between the two generating systems is shown in Fig. 2 in the part of the figure appearing between dashed lines. The power signals are clearly indicated as signal pairs. Each generating system exhibits two main feedback paths: one to regulate terminal voltage and the other to regulate speed (or frequency). The cross-feedback paths used to regulate the division of active and reactive loads between the two generators are also shown.

Although detailed derivations of the block diagram representations for the various components of an aircraft generating system have been previously reported in the literature,[1-3] it is appropriate here to summarize some of the basic features characterizing these component representations.

Alternator Representation

Because the alternator is the most complex component of the generating system, its block diagram is reproduced in Fig.

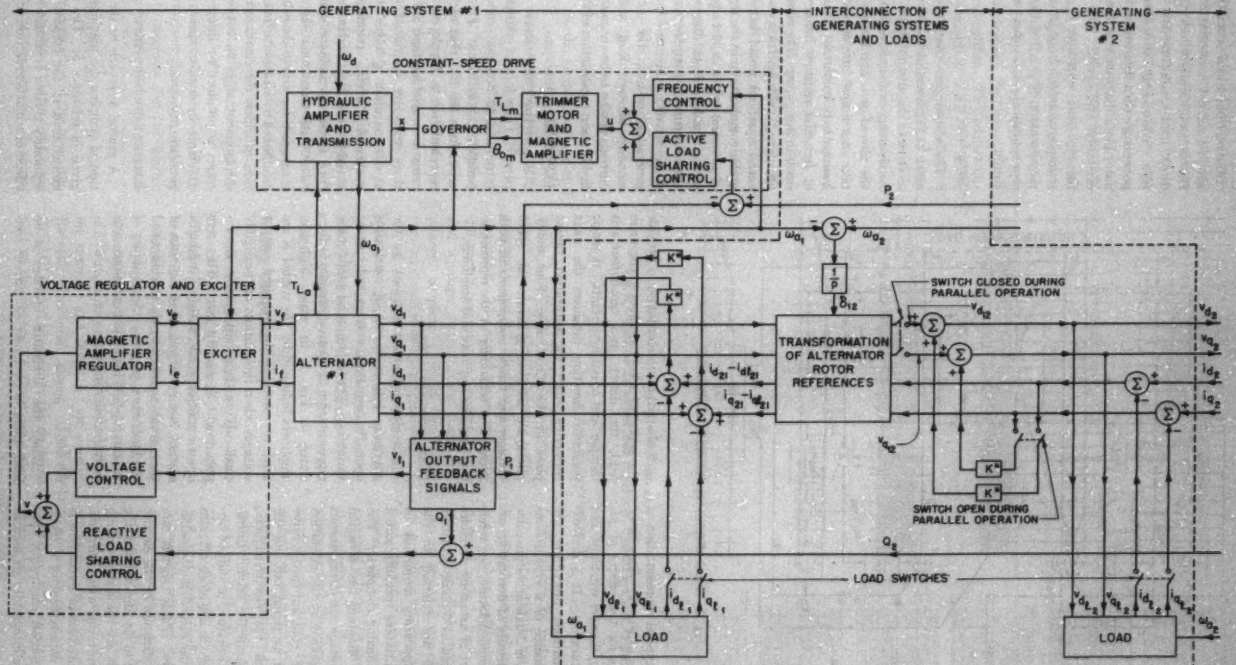


Fig. 2. Block diagram representation of two parallel aircraft a-c generating systems

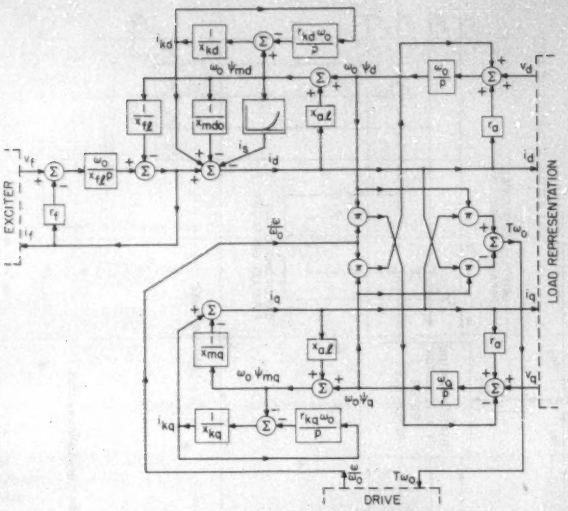


Fig. 3. Block diagram representation of alternator

3 to illustrate a typical representation of one of the blocks in Fig. 2. This block diagram accounts for the effects of saliency, saturation, damper windings, and variable-speed operation. It is derived[4] from Park's equations describing the transient performance of a synchronous machine in terms of 2-axis armature components. (For details of derivation, assumptions, and nomenclature, see reference 4.) This set of armature variables may be interpreted as 3-phase armature quantities which are referred to the rotor, thereby appearing essentially as equivalent d-c variables specially suited for use with an analog computer. The alternator block diagram has as signal inputs the field voltage and instantaneous angular speed supplied by the excitation system and the constant-speed drive respectively, and as signal outputs the field current and the electromagnetic torque required from the drive. Corresponding to the armature or output terminals of the machine are block diagram inputs expressed as direct- and quadrature-axis voltages. These signals are furnished by the diagram of the electric system connected to the alternator, while the output signals are the corresponding direct- and quadrature-axis currents. This particular direction of signal flow is imposed by the requirement of a block diagram representation employing only integrators. Five integrators are used in each alternator representation to provide the action of energy storages in the main field, the direct- and quadrature-axis damper and armature circuits. A single nonlinear function produces the main corrective effect for saturation taking place in the direct axis. Four multipliers are used to generate the speed voltages and torque components in both axes. The remainder of the representation employs summers and amplifiers.

Load Representation

The electric system supplied by the alternator imposes a set of constraints upon the armature voltage and current variables. The technique adopted for the representation of these constraints is illustrated in Fig. 2, in which high-gain amplifiers indicated by the symbol K are used to establish the required output voltages when their inputs are negligibly small. Were their gain infinity, the inputs would be exactly zero. This zero input condition would satisfy Kirchhoff's current law specifying that the sum of the currents supplied by the machines and the loads must be zero. As examples of various types of load representations, consider the following cases:

No Load

When the alternator operates at no load, there is obviously no current contribution from a load or another machine and the current constraint mentioned in the foregoing simply reduces to setting the direct- and quadrature-axis currents to zero. The no-load voltage terminal components are then determined by the alternator block diagram as a function of its excitation and speed. The removal of the high-gain amplifiers K corresponds to imposing zero voltage input signals to the alternator block; this, in effect, is the representation of a short-circuit constraint.

Static Load

The volt-ampere relations of the load must be expressed in the same variables as those of the alternator, i.e., in terms of 2-axis components referred to the rotor of the alternator. The load relations will take on a simple form only if the load is assumed to be balanced. The assumption of inductive-resistive balanced loads is made in this paper. The consideration of unbalanced loads with time-varying differential equations would have introduced undue complications in the representation of a system which is already very complex. The load block diagram furnishes the load current components in response to given voltage inputs. In an isolated generating system, the constraint of equality between load and generated currents is seen to be automatically satisfied in the representation of Fig. 2.

Paralleling

When two alternators operate in parallel, a new form of constraint must be introduced to account properly for the paralleling conditions. This constraint arises from the fact that each alternator representation is developed in terms of armature variables that are referred to their respective rotors. To be able to equate terminal voltages and to add generated currents requires that the variables be all expressed in the same reference frame (co-ordinate system). This co-ordinate transformation of variables from one rotor axis to the other is readily established by means of a "resolver" operation in which the characteristic relative displacement angle δ_{12} between the two rotors is obtained by integration of the difference in angular speeds of the two generating systems.

Exciter Representation

A general approach to the representation of complex d-c machine systems is given in reference 5. The aircraft d-c exciter used in the present system is a separately excited generator in which the effects of armature reaction and saturation

may justifiably be neglected for most conditions of operations. Its block diagram include a single integrator to account for the field storage and a multiplier to adjust the gain if the speed changes are very large.

Voltage Regulator Representation

The voltage regulator in this system consists of a 3-phase full-wave magnetic amplifier feeding the field of the exciter. The control circuit of this amplifier is fed by a single-phase full-wave bridge magnetic preamplifier. The control circuit of the preamplifier is fed by a bridge which compares the 3-phase full-wave rectified terminal voltage with the voltage of a voltage regulator tube. In addition, the control circuit of the preamplifier is fed by compensating circuits from the exciter and second-amplifier stage. The power supply for both magnetic amplifiers is the alternator output voltage and, in addition, the second stage is fed by current transformers from load current to supply excitation under fault conditions.

An analysis of the first stage yields a block diagram that contains one integration to account for the time constant caused by internal feedback in a full-wave magnetic amplifier. Also included in this block diagram is a saturation curve to account for the limited flux change in the saturable reactor cores. Since a full-wave magnetic amplifier responds to average values, the exciter feedback circuit (the only compensation having any appreciable effect) is represented by its d-c characteristics.

The 3-phase second stage has been analyzed with a difference equation approach.[7] Because of the complexity of the 3-phase circuit, it is found that there are three separate modes of operation and in each mode the behavior is described by fourth-order, nonlinear difference equations. By making reasonable assumptions, the difference equation in each mode may be reduced to a first-order linear difference equation, and this result is approximated by a differential equation in each mode. It is found that the first-order time constant thus obtained is different in each mode, but the difference is small. Therefore, the representation is reduced to a single time constant requiring one integrator, with one saturation to account for the limited flux change possible with the saturable reactors.

Drive Representation

A detailed derivation of the block diagram representation of the constant-speed drive is given in reference 8. As shown

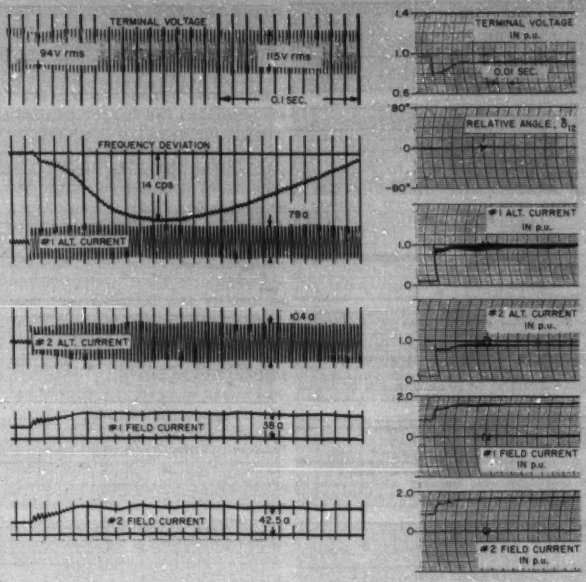


Fig. 4. Parallel alternators load switching: 0 to 60 kva at 0.75 power factor

A—Hardware test results
B—Computer test results

in Fig. 2, its inputs are the variable engine speed ω_d and the electromagnetic torque required to supply to electrical load on the alternator. The actual torque is slightly different if generator and exciter friction and windage, drive friction, shaft compliance, coupling backlash, and overrunning clutch effects are considered.

Two feedback corrective signals also appear as inputs to the drive block diagram: one is the output speed or frequency, the other is the difference $P_2 - P_1$ of real powers developed by the two generating systems.

Analog Computer Setup

Once the block diagram representation of the complete parallel generating system is obtained, it can readily be converted into an analog computer diagram with the introduction of suitable scaling.

To simulate the behavior of the drives, alternators, exciters, and regulators, a total of 33 integrators and 66 amplifiers (including summers and inverters) are needed. Reeves' electronic analog computer (REAC) is used to simulate the electrical part of the system. A Goodyear's electronic differential analyzer (GEDA) computer installation is used simultaneously to simulate the mechanical drives. The use of these particular computers was dictated solely by considerations of availability. Six diode-type function generators are required, two of them to account for generator saturation and four to perform the square root of the sum of the squares of the various components of terminal voltage and phase current for the two alternators. Fourteen servomultipliers are used to compute such quantities as torque, real power, reactive power, the square of terminal voltage and phase current, and various products of speed with other variables.

The time base used for the simulation is 500 seconds per second. Despite the fact that several minutes are required to complete a record of the frequency (or speed) transients, a long time base has been chosen for the following reasons: First, the large number of interlinked feedback loops required in the simulation of all the system equations produces a tendency to propagate any noise that may be generated in the computer. The lower loop gains which accompany the long time base

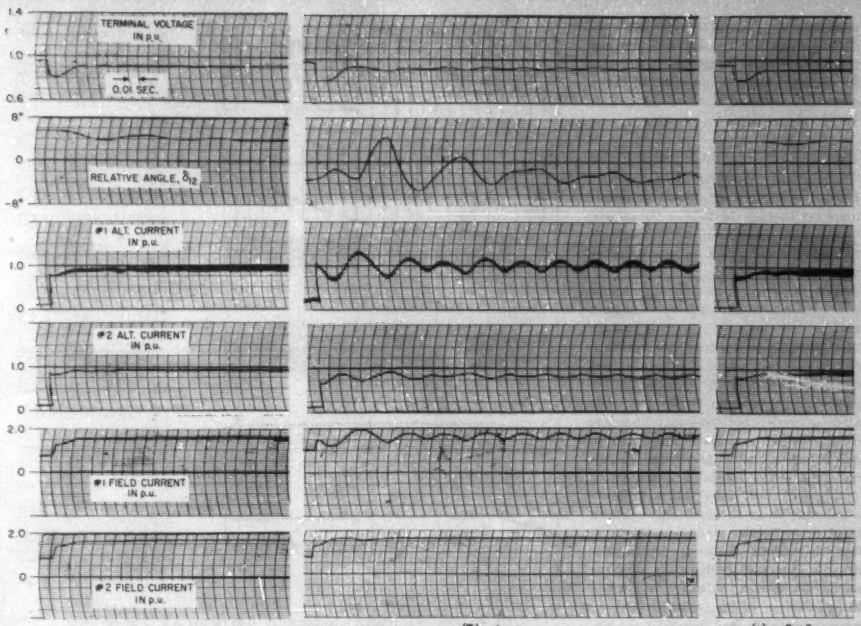


Fig. 5. Effect of sensitivities of load division circuits upon parallel alternator performance during load switching: 0 to 60 kva at 0.75 power factor

essentially lessen the disturbing effects created by noise. The second reason for having a long time base is that the computer servos can then follow the 400-cps components encountered in the various electrical variables (such as v_d and v_q). These components occur whenever a d-c offset transient term in the actual system is to be simulated. The 400-cps terms as referred to the computer time correspond to 0.8 cps signals which taxed the servos available whenever the amplitude was large. The use of a smaller time base aggravates this problem. Scaling is carried out so as to accommodate load-on and load-off transients and yield maximum steady-state accuracy.

Test Results

The best justification for the validity of the dynamic representations given in this paper is provided by a comparison of computer tests with actual equipment tests. A typical comparison of computer recordings and oscillographic data is illustrated in Fig. 4. The test performed on both hardware and computer consists in suddenly applying a load of 60 kva at 0.75 power factor on the two alternators operating in parallel at no load. The ensuing load-switching transients are obtained in the form of traces of terminal voltage, armature current, and field current for each alternator. The computer data for the alternators is expressed in a per-unit system having the following base values:

Armature phase voltage	120 volts
Armature phase current	111 amperes
Field current	23 amperes
Field voltage	1,740 volts
Speed	6,000 rpm
Power	40 kw

The agreement between computer and hardware tests is indeed remarkable if one considers the complexity of the system and the simplifying assumptions upon which the dynamic representations of the system components are based.

It should be remarked that the computer data for the alternator voltage and current are expressed in terms of the square root of the sum of the squares of the corresponding direct- and quadrature-axis components. The 2-axis formulation essentially eliminates the 400-cps component as a carrier signal from the computer traces. The variable amplitude 400-cps signal noticeable in the alternator traces is in the nature of noise.

The main conclusion to be drawn from an examination of the hardware and computer tests is a clear demonstration of the ability of the analog computer to simulate transient conditions accurately. The computer can be used with confidence to determine whether a generating system will meet given conditions such as military specifications regarding recovery time and voltage transient limits. One of the basic advantages of such a computer simulation is the ease with which the effects of design changes may be assessed and improvement in system behavior finally achieved.

The effects of varying the load division sensitivities (both active and reactive) upon the parallel operation of the system were examined in order to illustrate the use of the computer for design studies. The procedure consisted in modifying the values of the gains of the load division circuits and in determining the corre-

sponding improvement or deterioration of the transient response of the system under various conditions of load switching. A situation of particular significance is illustrated in Fig. 5, which clearly displays the instability of the system arising from a large value of reactive load division sensitivity. The transients are initiated by applying 60 kva at 0.75 power factor on an unloaded parallel system. Fig. 4 shows the conditions for which the sensitivities have their normal values. Fig. 5(A) shows the effect of first doubling the sensitivity of the reactive load division, then (B) changing this sensitivity by a factor of eight from its normal value, and finally (C) doubling the sensitivity of the active load division circuit with the reactive load division circuit sensitivity at its normal value.

It is worth noting from Fig. 5 that the oscillations in armature currents at a frequency of about 18 cps are essentially 180 degrees out of phase for the two alternators. A similar frequency of oscillation may be observed in the field current of alternator 1 only. The reason for the difference in behavior of the two alternators is that the reactive load division corrective signal is applied to the voltage regulator of only one system (alternator 1). A further examination of Fig. 5 discloses the effect of the reactive sensitivity upon the mechanical part of the system as illustrated by the oscillations in angular displacement δ_{12} between the two generating systems. This situation is a clear indication of an existing interaction between active and reactive load division circuits.

Conclusions

The excellent agreement between hardware and computer test results provides the justification for the general approach used in establishing the simulation. Although only particular configurations (two parallel a-c generating systems with static balanced 3-phase loads) and specific operating conditions have been examined, the techniques of simulation presented in this paper can be readily extended to cover new situations such as multimachine systems with induction and synchronous motor loads or with unbalanced loads. The computer study described in the paper should be interpreted as an illustration of the kind of problems that can be tackled by means of analog computation. The adoption of analog computation for design and synthesis prior to construction of the actual equipment combination to be used is yet to be explored more completely. Typical of the studies that can be treated effectively after arriving at a satisfactory simulation of an aircraft electrical system are the following:

1. Study of the problem of paralleling transients.
2. Study of the interaction of voltage and speed regulators.
3. Study of voltage and frequency modulation and load instability problems.
4. Study of the effect of the damper windings on the dynamic behavior of the parallel system.
5. Study of new techniques for improving the stability of the over-all system; in particular, the manipulation of various types of feedback such as, to mention one example, the use of corrective signals proportional to the rate of change of displacement angle δ_{12}.
6. Study of other system configurations obtained by replacing a system component with another of different design or by modifying the design of a given component.

References

1. Analog Computer Representations of Aircraft Parallel Alternating-Current Generating Systems, L. V. Boffi, M. Riaz, P. E. Smith, Jr. *WADC Technical Note 56-384*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Sept. 1956.
2. A Performance Study of an Aircraft A-C Generating System Using Analog Computation, M. Riaz, P. E. Smith, Jr. *WADC Technical Note 57-160, Ibid.*, Sept. 1957.
3. Final Report—Study of Aircraft Alternating-Current Electrical Power Generating Systems, M. Riaz, D. C. White. *WADC Technical Report 57-598, Ibid.*, Oct. 1957.
4. Analog Computer Representation of Alternators for Parallel Operations, Luiz V. Boffi, Vinton B. Haas, Jr. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 153–58.
5. Dynamics of D-C Machine Systems, M. Riaz. *Ibid.*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 365–70.
6. Full-Wave Bridge Magnetic Amplifiers with Inductive Loads, H. H. Woodson. *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 7–9.
7. A Representation for a Three-Phase Magnetic Amplifier, L. Maisel. *WADC Technical Note 57-158*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Aug. 1957.
8. Dynamic Representation of a Hydraulic Constant-Speed Drive, Paul E. Smith, Jr. *AIEE Transactions*, vol. 77, pt. II, Mar. 1958, pp. 28–34.

Tramp-Iron Remover for a Field Forage Harvester

F. C. JACOB
NONMEMBER AIEE

J. B. DOBIE
NONMEMBER AIEE

An electronic tramp-iron remover for chopped hay in pneumatic conveyors has been previously described for application in fixed installations such as barns or feed-mills.[1,2] When green chopped hay is fed, however, a field harvester is the only machine that handles the hay pneumatically. The tramp-iron remover can be adapted for use on a field harvester to make a self-contained unit that is in some ways simpler than the one described earlier for fixed locations.

A tramp-iron remover on a field harvester does the same job as one in a fixed location. It removes nails, wire, and other metal objects likely to puncture the stomach of a cow. Since these are usually iron or steel, only ferrous metals are considered. Pieces smaller than the equivalent of a half-inch length of baling wire are not considered hazardous.

Conditions to Be Considered in Design

The following conditions must be considered in adapting the tramp-iron remover to a field forage harvester:

1. The blower pipe is relatively short.
2. The chopping machinery operates through a wider speed range than is usual for a fixed installation.
3. If an electrical rejection system is used, the source of power is a battery or generator. The battery may change output with time or weather; the generator, with speed.
4. The combined vibration from a gasoline engine on the field machine and from operation over rough fields may be greater than vibration conditions at a fixed location.
5. The condition of the hay may be more variable when it comes directly from the ground than when it has been naturally mixed by bulk handling or cured in the bale.
6. To prevent recontamination, the rejected hay and iron should not be cast back onto the hay field.

Paper 58-1006, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted April 16, 1958; made available for printing June 12, 1958.

F. C. Jacob and J. B. Dobie are with the University of California, Davis, Calif.

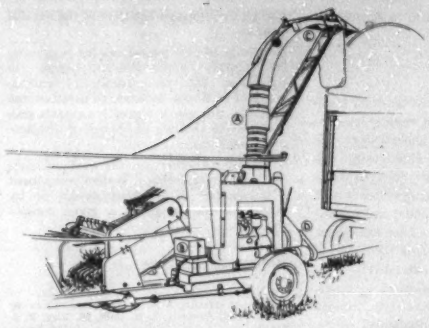
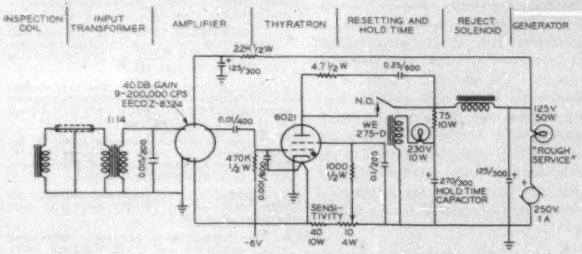


Fig. 1 (left). Field forage harvester with tramp-iron remover

A—Inspection coil
B—Amplifier
C—Reject gate
D—Generator

Fig. 2 (below). Schematic diagram of electrical system



Description

The over-all scheme is shown in Fig. 1. Tramp iron moving with the hay generates a voltage in the inspection coil *A*. The amplifier *B* receives the signal from the coil and energizes the reject gate *C*. A small amount of hay is diverted into the bag along with the tramp iron. Power is supplied to the tramp-iron remover by the generator *D*.

Design Criteria

Inspection Coil

The same criteria that govern the design of an inspection coil for a stationary installation also govern that for a mobile one. In fact, the coil assemblies are interchangeable. To achieve the desired sensitivity the magnetic field strength on the pipe axis should be 75 to 125 gauss, and the coil should have 5,000 to 10,000 equivalent turns (number of turns multiplied by the step-up ratio of the amplifier input transformer).

The impedance of the amplifier input transformer secondary is about 100,000 ohms. The wire size in the coil should be chosen so that its resistance reflected to the input transformer secondary will be small compared to 100,000 ohms.

The rise time of the signal is slower than 2 milliseconds. Spurious responses having faster rise time can be attenuated by capacitive loading of the inspection coil at the amplifier input. If the distributed capacitance of the coil is not sufficient, a lumped capacitance may be added at the amplifier to make the coil system parallel resonant at 200 to 300 cycles per second. In Fig. 2 a 0.015-microfarad capacitor serves this function.

The signal in the inspection coil is produced by a change in magnetic flux, and anything that causes such a change produces a signal. Spurious signals are produced if,

1. the pickup coil winding moves with respect to the magnetic field,
2. the magnetic flux changes because of structural distortion or because of a changing air gap in the magnetic path,
3. a magnetic or electrically conducting object, such as part of the forage harvester itself, moves in the field of the magnet.

The first two factors, movement of the pickup coil winding and structural factors connected with the magnetic flux path, are largely controlled by design and workmanship. All parts of the inspection coil assembly should be rigidly fixed with respect to one another. However, vibration transmitted from the harvester machinery can cause distortion in what may appear to be rigid parts. Tests with experimental models indicated a means for vibration isolation of the inspection coil was necessary. This was achieved by using rubber hose couplings between the harvester pipe and the short section of plastic pipe on which the inspection coil was mounted.

The third factor, nearness to moving machine parts, affects the placement of the inspection coil assembly on the harvester. Any machine part must be considered a moving part because of vibration, and must not be closer to the inspection coil than a distance equal to the blower pipe diameter. The part of the blower pipe that passes through the inspection coil assembly must therefore be replaced by plastic pipe. Even aluminum pipe is not satisfactory, since eddy currents would be induced in it which, in turn, would have a magnetic field associated with them. The plastic pipe should extend each way from the inspection coil for a distance at least equal to the diameter of the blower pipe. The use of small stainless steel or brass fasteners on the coil assembly does not cause appreciable interference.

The minimum length of the inspection magnet is one that provides a workably uniform magnetic field inside the pipe. In a nonuniform magnetic field, the detection sensitivity will vary. It will be least sensitive along the pipe axis and greatest along the pipe sides. The system should be designed to detect the smallest dangerous piece of tramp iron at the least sensitive point in the pipe; then smaller pieces will be detected in more sensitive areas. If the inspection coil assembly is made long to produce a uniform magnetic field, the required space and the cost become objectionable. Practical dimensions are obtained if the length of the magnet is equal to the pipe diameter.

Response Speed and Timing

The time required for tramp iron to move in the forage harvester blower pipe from the inspection position to the reject-gate position (82 inches) was measured, using two inspection coils as previously reported.[1] For 50 measurements the longest transit time was 182 milliseconds and the shortest was 45 milliseconds. Slower transit than 182 milliseconds could occasionally be expected. A design hold time of 250 milliseconds for the machine under normal load was adopted as one which would allow most of the tramp iron to be caught without undue loss of hay.

Transits faster than 45 milliseconds

probably would not occur with the forage harvester used, since this represents a velocity corresponding to the blower tip speed. A gate-opening time of no more than 40 milliseconds was selected as that necessary to catch the fast-moving pieces of tramp iron with a margin of safety.

To attain a gate-operating speed of 40 milliseconds or less it was necessary to use solenoids energized from a voltage source higher than that of the forage harvester storage battery. Since 5 amperes represented the maximum current-carrying capacity for a small thyratron and a high-speed relay, this current was selected as a maximum permissible value in the solenoid circuit. With this restriction, solenoids operating from 6, 12, or even 24 volts were too slow. A power supply of 250 volts was selected as being adequate and is a standard value for d-c generators.

Amplifier

The inspection coil output is a pulse of several tens of millivolts. The value of the signal depends on the size, orientation, and speed of the piece of iron. The amplifier must raise this signal level to operate the reject solenoid and must develop a sustained pulse to provide the hold time discussed in the foregoing.

It is necessary for the machine to extend the hold time if two or more pieces of tramp iron are in transit and separated by time intervals less than the hold time for the gate. In the extreme case, a long piece of wire might enter the chopper and be cut into pieces that would follow one another in rapid sequence over a period much longer than the hold time needed for a single object. A circuit for performing this hold-time extension can be seen in the wiring diagram of Fig. 2; this involves the thyratron resetting relay, the hold-time capacitor, and the reject-gate solenoids. The solenoids are energized by the capacitor charging current. The capacitor is discharged by the thyratron and the thyratron is reset by the relay. The effective hold time is the interval starting after the thyratron is reset and ending when the capacitor charging current has dropped to a value below that necessary to energize the solenoids. Passage of a piece of tramp iron through the inspection coil during the effective hold time again causes the thyratron to discharge the capacitor, initiating another hold-time period as an extension of the previous hold time.

Electrically, every hold cycle is the same. The energizing current to the solenoid and the solenoid torque are independent of the shape or magnitude of the inspection coil signal as long as the signal is sufficient to trigger the thyratron. The restoring torque on the reject gate produced by the air and hay movement, however, will be a function of blower speed and loading. The actual hold time can therefore be expected to vary under field conditions.

If the operation of the amplifier and reject-gate solenoids is dependent on a generator for operation, as it was in the experimental machine, then the problem of voltage variation must be considered. Under heavy load, the harvester engine may slow to 80% or less of full-load speed with consequent reduction of the generator output. The solenoids must therefore have some torque reserve at normal load condition. The amplifier, too, was designed to accommodate the change in generator output down to 80% speed. To design for a much greater range would involve added expense. If the engine slows to below 80% of normal, the chopper is on the verge of plugging, a condition which can be averted by an alert operator.

The Reject Gate

The reject gate is located as near the end of the blower pipe as possible and still obtain satisfactory deflection of the rejected hay into a suitable receptacle. The length and height of the blower pipe are limited, particularly if it is for use with a covered forage wagon. For the forage harvester shown in Fig. 1, the coil-to-gate distance was 82 inches, measured along the upper side of the pipe. Since the pipe is open along the bottom in the region of the reject gate, there is negligible static pressure, as contrasted to the high static pressure in the closed system needed for the stationary iron remover. This difference permits great simplification in gate design for the field harvester.

The curved section of blower pipe assures location of the hay stream at the upper side of the pipe, reducing the size of gate vane required to provide positive deflection of the rejected hay. A vane 5 inches long, operating through an angle of 45 degrees, is minimum. When made of aluminum sheet, the vane has a mass small enough that two small rotary solenoids fastened to the ends of the gate shaft are sufficient to move the vane to the reject position in less than 40 milliseconds. The solenoids used are limited to 45-degree rotation and hence serve to position the gate vane at both extremes. Since air and hay provide the force to return the vane as soon as the electrical circuit is disconnected, no return spring is required. The pipe and vane are rectangular to facilitate a close fit along the edges, thereby reducing the possibility of forage wedging around the vane and slowing its operation. The shaft can be recessed in a slot in the upper side of the pipe to provide a smooth path for the forage under normal operating conditions.

Disposal of Rejected Hay

In field trials, rejected material varied from zero to one cubic foot per acre. Two methods of catching the rejected hay and tramp iron were used.

1. A box fastened to the front of the forage wagon following the harvester. Since the box can be made large, it has the advantage of requiring infrequent attention and can be emptied when the wagon is out of the field. A box should be provided for each wagon.

2. A canvas bag attached to the blower pipe, as shown in Fig. 1. During operation in uneven fields the relative position of the bag and the reject gate does not change as it does with the wagon-mounted box. A bag that can be quickly detached or easily opened facilitates emptying at the edge of the field.

Performance

Catch trials have demonstrated removal efficiency of better than 90% of dangerous tramp iron, using the design criteria outlined in the foregoing. One-fourth to one-half pound of hay is carried with each rejection. Because of the possibility of a very long transit time, removal of all tramp iron could not be attained without rejecting an objectionable amount of hay. Reducing the hold time from the design point of 250 milliseconds to 150 milliseconds lowered the catch efficiency to about 65% and resulted in about a one-third saving in rejected hay.

Moisture content had no effect on performance of the tramp-iron remover for forage ranging from field-cured to green chopped hay.

Sharp chopper knives and a harvester engine in good running condition are important factors in obtaining optimum results with the tramp-iron remover. Together with an alert operator they provide assurance of nearly uniform conditions and steady operation.

References

1. Electronic Detection and Removal of Tramp Iron from Chopped Hay, John B. Dobie, et al. *Agricultural Engineering*, St. Joseph, Mich., vol. 34, no. 6, 1953, pp. 377–81.

2. Removing Tramp Iron from Chopped Hay, J. B. Dobie, F. C. Jacob. *Electronics*, New York, N. Y., vol. 27, no. 5, 1954, pp. 134–36.

Artificial Cooling of Livestock in Hot Climates

C. F. KELLY
NONMEMBER AIEE

T. E. BOND
NONMEMBER AIEE

N. R. ITTNER
NONMEMBER AIEE

THE possibilities of artificially increasing the comfort, or, more correctly, the efficiency of use of feed by farm livestock, are great. The range of optimum temperatures for laying hens is 55 to 65 F (degrees Fahrenheit), for dairy cows 50 to 60 F, and for swine 65 to 75 F. Size and age of animal, sex, coat, etc., are all related to its ability to resist high or low temperatures. Adverse effects are apparent sooner from high than from low temperatures; in general, livestock are considered to be depressed by temperatures over 75 F.

In Fig. 1 it can be seen that the isotherm representing the July average temperature of 75 F divides the United States into two regions. South of this line, July temperatures will always average about 75 F, and north of it below 75 F. The bar charts show that capacity for generating electricity and production of electricity in the warm area are slightly less than half the total U.S. capacity and production. This area contains slightly more than 50% of the cattle raised in the United States; its farm value is somewhat less than half the U.S. total.

It is interesting to note that the 75 F July average isotherm closely follows the line south of which, according to one manufacturer, air conditioning of houses is required for 1,400 to 2,900 hours per cooling season.[1] This estimate is based on outside temperatures of 80 F and higher.

In California's Imperial Valley, where the co-operative investigations to be described were made, the climate is very well suited to hot-weather studies with livestock. The mean monthly temperature is above 75 F for about 6 months of the year, exceeds 85 degrees for nearly 4 months, and reaches 90 degrees for 2 months. Daily maximum temperatures exceed these levels considerably. The average weather conditions for the years of the tests are shown in Table I.

Electric power can be used to cool livestock in several ways. Among these are: cooling the drinking water to cool the animal internally; increasing air motion or decreasing ambient air temperature to increase convection cooling; and providing a cooled slab for the animals to lie on to increase the rate of heat loss by conduction. Each of these methods must be used in connection with well-designed structures that will either protect the animals from unfavorable natural conditions or exploit favorable conditions. (This phase of University of California and U.S. Department of Agriculture research—the design of shades, exploitation of the heat sink of the north sky, the construction of corrals to use natural air movement—has been described before.[2-7]) The studies using electric power for cooling are described herein.

Cooled Drinking Water

In many desert areas, livestock drinking water comes from irrigation ditches or from exposed uninsulated water pipes, and its temperature at the drinkers is 90 F or higher. Lowering this temperature seemed to be one way of increasing gains. This could be accomplished with a concentrated unit, installed near the water troughs, and the animals would not be restricted in their movement. In the case of cooling by low ambient temperature or increased air velocity, the animals must remain at one location to receive the beneficial effects.

In 1950 a study of the effects of cooled drinking water upon beef cattle was initiated at the University of California's Imperial Valley Field Station near El Centro, Calif. The water temperature was lowered by both refrigeration and evaporation. In the case of refrigeration, the water temperature was usually lowered to 65 F, the ambient air temperature considered to be "comfortable" for most classes of livestock. With the evaporation method the water temperature was lowered to within 2 or 3 degrees of the air wet-bulb temperature.

The mechanical refrigeration system for cooling the water consisted of a 3-hp (horsepower) compressor with direct expansion coils of 60 feet of 5/8-inch copper tubing immersed directly in each drinker; see Fig. 2. The drinkers were sections of 36-inch-diameter concrete tile, fitted with a concrete bottom and a wooden top and having an opening large enough for only one animal to drink at a time. The sides of the tank were insulated with 2-inch mineral wool batts. Water level was maintained constant by a float. This system easily maintained the water temperature at 65 F for 20 head of cattle even in the hottest weather, when the initial water temperature was 90 F.

The evaporation cooling system, shown in Fig. 3, was a commercial air-conditioning condenser cooler tower having 12 36- by 36-inch trays of 1/4-inch mesh over which water was pumped at the rate of 10 gallons per minute by a 1/2-hp centrifugal pump. The insulated supply and storage tank for the cool water, 4 by 4 by 4 feet, immediately beneath the tower, had a capacity of 300 gallons, enough for 15 or 20 head of beef cattle for one day. The animals obtained their

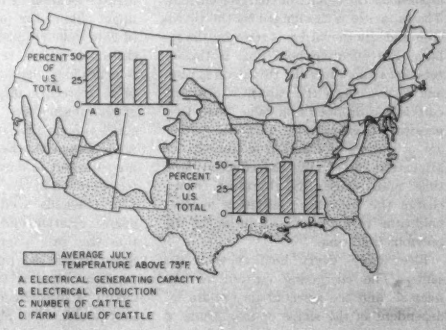


Fig. 1. Division of United States into two areas by July 75 F average temperature isotherm, and relation of electrical generating capacity and production, and number and value of cattle, in two areas

Paper 58-1007, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19-22, 1958. Manuscript submitted April 29, 1958; made available for printing June 12, 1958.

C. F. KELLY is with the University of California, Davis, Calif.; T. E. BOND is with the U.S. Department of Agriculture, Davis, Calif.; and N. R. ITTNER (deceased) was with the University of California, El Centro, Calif.

Table I. Weather During Feeding Periods

Data for Imperial Valley Field Station During Test Periods, 1948 to 1957 Inclusive

Year	Period of Feeding Trial	Air Temperature, F			Air Relative Humidity, %			
		Average Maximum	Average Minimum	Mean	Average Maximum	Average Minimum	Noon	Mean*
1948	July 28–Sept. 20	108.1	74.7	91.4	55	17	25	33
1949	July 12–Sept. 16	104.3	73.4	88.9	71	24	31	44
1950	June 21–Sept. 9	105.1	74.8	89.9	76	28	39	47
1951	July 3–Sept. 27	103.7	71.8	87.7	78	29	36	50
1952	July 2–Sept. 11	100.7	77.7	91.8	66	26	26	40
1953	June 30–Sept. 22	105.5	77.1	91.3	61	19	24	36
1954	June 24–Sept. 16	103.6	76.4	90.0	67	19	28	39
1955	July 6–Sept. 14	103.2	78.3	90.3	80	23	29	46
1956	June 27–Sept. 5	102.4	74.2	90.3	68	17	22	36
1957	June 26–Sept. 18	107.4	74.0	90.6				

* The mean relative humidity is the average of all 2-hour readings throughout the period.

drinking water as needed from two drinking cups attached to the tank near the bottom. The pump and circulating system were operated from 2 to 5 a.m. each night during the period of lowest dew point, being turned off and on by an electric timer. No water was added during the daytime; the clock, by means of a solenoid valve, also filled the tank with warm water at the time the circulating pump started. Three hours of circulation usually reduced the water temperature to within 2 or 3 degrees of wet-bulb temperature, which at the coolest part of the night ranged from 55 to 75 F. The temperature of the water at the bottom of the tank did not increase more than a few degrees during the day.

The results of 6 years of experiments with cooled drinking water are summarized in Table II. The following general conclusions can be drawn:

1. Average daily gains per animal were increased by about 0.30 lb (pound).
2. Under the conditions of the test, there was little difference in results between water temperatures of 60 and 70 F.
3. Salt in excess added to the feed increased consumption of cold water but did not increase daily weight gains.
4. Water consumption decreased slightly when the water was cooled.
5. Cooling by evaporation did not lower water temperature enough to produce increased gains.

Fans

In an earlier study a comparison of wire or cable corrals with wood plank corrals had shown gains to be considerably greater in the wire corrals.[7] In analyzing the data for the cause it was concluded that while radiation environment in the wire corrals was considerably lower than in the wooden ones, the greater natural breeze through the more open wire corrals must also have increased heat loss by convection. To separate these two variables (radiation and convection) as much as possible in field trials, a series of studies using fans was started in 1955. These are summarized in Table III.

Low-speed ventilation fans, 42 or 44 inches in diameter and rated at about 17,000 cfm (cubic feet per minute) free air delivery, were used in all the tests. They were mounted on the fence line several feet above the ground so as to blow air lengthwise of the shades and with the prevailing breezes. A single fan in one of the experimental pens at El Centro and a string of four fans in a commercial feed lot at Blythe, Calif., are shown in Figs. 4 and 5. The design and results of this test are summarized in Table III.

The effect of the fans upon weight gains in wood corrals was startling. In 1955 the average daily gain of the "fanned" animals was 44% more than that of the "unfanned" animals; in 1956 it was 22% more. These gains were accomplished with 30% and 19% less feed per 100-lb gain, respectively. The difference between the results for the 2 years was due to better gains for the controls in 1956 rather than to decreased gains in the fanned animals. The generally cooler summer of 1956, with lower relative humidity, may have been responsible.

In 1957 the results with fans, both at the field station and at a co-operator's commercial feed lot in Blythe, were not so marked. All fans were installed in wire corrals, and the difference in air velocity between fanned and unfanned

Fig. 2 (left). Expansion coils for refrigerated drinking water

Fig. 3 (right). Evaporation tower for cooling drinking water

Table II. Weight Gains of Beef Cattle with Cold Drinking Water Versus Warm Drinking Water

Year	No. of Animals Each Lot	Cooled Water Temperature, F	Initial Weight, Lbs	Average Daily Gain, Lbs		Feed Per 100 Lbs Gain, Lbs	
				Cooled Water	Controls	Cooled Water	Controls
1950*	4	65	841	1.46	1.07	1,618	2,092
1951†	5	75	842	1.56	1.58	1,366	1,371
1952‡	10	65	764	1.04	0.78	2,221	2,480
1953§	7	65	875	2.01	1.57	1,073	1,279
1953‖	6	65	868	1.51	1.18	1,393	1,094
1954¶	7	65	815	1.70	1.51	1,030	1,157
1954**	7	63	814	1.50	1.47	1,258	1,233
1954††	7	71	815	2.03	1.94	860	867
1955‡‡	7	90	668	1.69	1.29	1,121	1,330
1955‡‡	7	70	670	1.79	1.29	1,089	1,330

* Ration was alfalfa and barley or oat hay.
† Water cooled by evaporation, temperature ranging between 66 and 80 F, depending upon air humidity. Ration was hay and grain.
‡ Ration, alfalfa hay.
§ Ration, hay and grain.
‖ Ration, alfalfa and barley or oat hay.
¶ Pelleted ration of barley, alfalfa, and oat hay, and dried beet pulp.
** Pelleted ration as above plus 10% salt to increase water intake.
†† Water cooled by evaporation to 50–79 F. Control water temperature averaged 81.4 F. Pelleted ration of barley, hay, and dried beet pulp.
‡‡ Ration, hay and grain. Same controls for both experimental groups.

Table III. Comparison of Weight Gains of Hereford Beef Cattle with and without Fans

Year	No. of Animals on Test	Average Initial Weight, Lbs	Average Daily Gain, Lbs		Feed Per 100 Lbs Gain, Lbs	
			Fanned	Control	Fanned	Control
1955*	7	669	2.32	1.29	924	1,330
1956†	7	651	2.40	1.87	798	982
1957‡	6	642	2.62	2.50	811	816
1957§	6	644	2.52	2.50	868	816
1957‖	6	648	2.25	2.50	912	816
1957¶	251	415	1.45	1.39	661	731

* Average air velocity at center of fanned pen was 3.7 miles per hour, in control pen 0.63 mile per hour. Wood corrals.
† No continuous velocity measurements. Wood corrals.
‡ Fan on 12 hours at night. Wire corrals.
§ Fan on 12 hours during day. Wire corrals.
‖ Fan on continuously. Wire corrals.
¶ Fan on 9 a.m. to 5 p.m. Wire corrals. This test on feed lot of Chaffin Ranch (Ben Olsen, manager), Blythe, Calif.

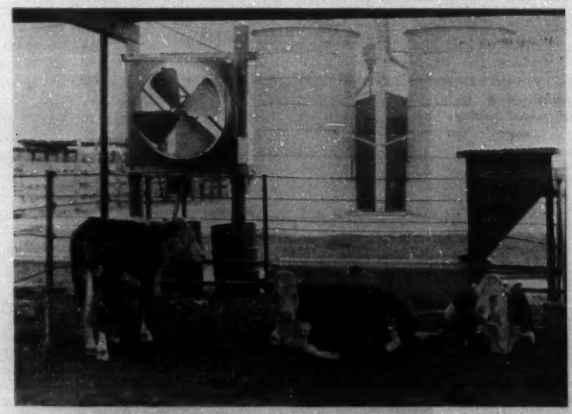


Fig. 4. Fan in experimental pen, El Centro, Calif.

pens was not so great. Furthermore, the natural summer air velocity at both El Centro and Blythe was greater than normal, and, as stated before, wire corrals also have the advantage of greater radiation cooling, especially if surrounded by a green growing crop.

The use of fans in open pens presents a problem even greater than their use to protect citrus orchards from frost; the period of operation is much longer, and the fans must be located so that the animals can be constantly in its breeze, even when under a shade with a moving shadow. The principle of heat transfer also differs; the orchard fan mixes warm air with cold to prevent freezing, whereas in the corral the purpose of the fan is to increase the transfer of heat from the animal's surface to the ambient air. (It is interesting to note that the ambient air temperature in the two fanned wood corrals in 1955 and 1956 was 2 or 3 degrees higher than in the control corrals.)

Selection of a fan for this purpose is governed by several considerations: (1) The fan must be low in initial cost and economical in its use of power. (2) It must be relatively rugged and able to operate under hot and dusty conditions. (3) It must produce a pattern of air movement that is suitable for the size and shape of corral and the shades used.

The tests discussed herein were made with low-cost ventilating fans of the propeller type, designed to move air against low pressures. Because of their economical construction (sheet metal blades, a strong but lightweight frame, and V-belt drives), advantage can be taken of the fact that, for a given thrust, increasing the fan diameter reduces power requirements by the same ratio. If the diameter is doubled, the power requirement is cut in half. If a greater capacity is required, fan size rather than fan speed may be economically increased, as power requirements increase directly with fan diameter but with the cube of the fan speed. At the present time 42-inch-diameter fans, capacity 15,000 to 17,000 cfm, complete with a 3/4-hp motor, can be obtained for approximately $200.

Probably the greatest problem in applying fans to open corrals or other structures is obtaining uniform distribution of velocity over the entire area occupied by the animals. Fig. 6 shows the results of observations of air velocity at a level 3 feet above the ground. At this time the natural air movement was between 25 and 50 feet per minute in the same direction the fan was blowing. The spread of the fan jet is shown to be

in accordance with the observation of Brooks et al[3] that the outline of the jet is a cylindrical cone with the propeller located in the cone three diameters from the apex, making an included angle of about 24 degrees.

Several angles of downward tilt of the fan axis were checked. The data in Fig. 6 are for an angle of 7 degrees, which is believed correct for a spacing of 50 to 75 feet between fans of this capacity.

The fans must always be supported at a level high enough above ground so that the air jet will be in a general downward direction on the backs of the animals and so that there will be no danger from the blades. The test fans were always positioned with the fan center line at least 8 feet above the ground.

The first increment of increased cooling by forced convection will be obtained with less cost than further increases because of the relationship between air velocity and surface film conductance. As the rate of heat loss increases with approximately the 0.5 or 0.6 power of the air velocity, it is evident that, with a given fan size and air temperature and humidity, doubling the air velocity for instance, will only increase the rate of heat loss by 40% while it may increase power consumption by 700%, assuming that air velocity at a given point varies directly with capacity of fan.

Air Conditioning

Most investigators of the effect of environment upon livestock agree that ambient air temperatures between 60 to 70 F are favorable. In order to take advantage of the low summer humidities in the Imperial Valley, an evaporative or "desert" cooler was used in a 1948 test to lower the temperature of the air. A cooler with a rated capacity of 2,500 cfm installed in the south end of a 13- by 22-foot 3-sided insulated shelter. The north end, opposite the cooler, was open to the corral. The cooler was operated from 9 a.m. until 5 p.m., providing about one air change per minute.

Five Herefords, with an average initial weight of 429 lbs, and two Brafords were housed in this shelter for a 54-day feeding trial, July 29 to September 20, and compared with similar animals having a shade but no cooler. The average daily gain of the cooled Herefords was 1.05 lbs, and of the uncooled 0.69 lb. The Brafords, being more tolerant to heat, showed no significant difference in gains.

The cooler increased animal comfort both by lowering air temperature and increasing air velocity. The dry-bulb

Fig. 5. Fans in commercial feed lot, Chaffin Ranch, Blythe, Calif.

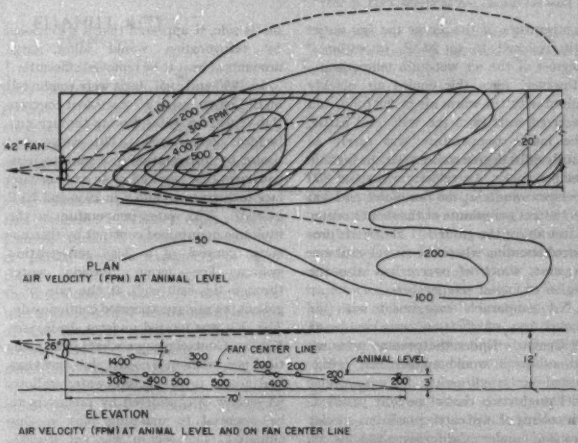


Fig. 6 (above). Distribution of air by low-speed fan

Fig. 7(right). Pigs on 65 F concrete slab

Table IV. Comparison of Weight Gains of Swine with Cooled Slab and Wallow

Year	No. of Pigs Each Lot	Average Slab Temperature, F	Initial Weight	Average Daily Gain, Lbs		Feed per 100 Lbs Gain, Lbs	
				Cooled Slab	Control	Cooled	Control
1956	7	65	52	1.37	1.52	349	360
1957	5	70	82	1.61	1.37	385	421

Table V. Average Electric Power Used per Summer Season, per Day, and per Animal*

Type of Cooling	No. of Years in Tests	Power, Kwhr		
		Total for Test	Per Day	Per Animal Day
Desert cooler	1	88	1.63	0.230
Cooled drinking water				
Refrigeration	4	889	12.7	0.814
Evaporative tower	2	105	1.50	0.291
Fans in corral†	3	1,090	15.6	2.216
Cooled slab for swine				
For refrigeration		1,029	14.6	2.44
For water pump		334	4.7	0.84
Total	2	1,363	19.3	3.28

* Based on average number of days in feeding trials and actual number of animals on test.
† Does not include test at Blythe, Calif, in 1957.

temperature of the air at the fan outlet was reduced 15 to 20 F, to within 5 degrees of the air wet-bulb temperature. However, since this cooled air quickly mixed with the warm air of the room and received heat from the walls and cattle, the inside temperature was rarely as much as 10 degree cooler than the outside temperature. Air velocity was about 850 feet per minute at the fan outlet and 300 to 500 feet per minute at the shelter center 3 feet above the ground. The cattle preferred standing where the air velocity was greatest, which of course was also the region of lowest temperatures.

No comparable experiment with air cooled by refrigeration were made at El Centro. Under the present price relationships it would appear that refrigerated air conditioning for commercial beef production cannot be fully justified. Air cooling of beef cattle presents a special problem under California methods of management, for the structure must be open enough for free in-and-out movement of the animals. However, with the development of better filters and heat exchangers for air conditioners and with further research in low-cost large-area shelters, it is possible that air conditioning may become practical.

Cooled Slab

Standing livestock lose little heat by conduction because of the small area of their feet or hooves. Beef cattle spend little time lying down. Swine, on the other hand, have been observed to lie down as much as 22 hours per day during hot weather. As approximately 20% of the hog's surface area comes in contact with the floor when the animal is lying on its side, it appeared that a slab cooled by refrigeration would allow large amounts of heat to be removed efficiently.[9]

In 1956 and 1957, tests were conducted at El Centro, using a 7- by 9-foot concrete slab cooled by water flowing through embedded copper coils. The water was cooled by refrigeration in a tank to temperatures that resulted in an average surface temperature of 65 F in 1956 and 70 F in 1957. The water temperature in the tank was maintained constant by thermostatic control of a 3-hp refrigeration system. The pump, forcing water through the slab coils at the rate of 7 gallons per minute, operated continuously. The slab was located under a shade; see Fig. 7. Control pigs were kept in an identical pen, without a cooled slab, but they had access to a concrete water wallow. (A wallow is considered by producers to be essential to maintain pigs in this climate.) The wallow was not shaded, but shade was provided nearby.

The results of the 2 years' tests are shown in Table IV. In 1956 the pigs on the slab did not gain as rapidly as the control pigs. In 1957, however, the slab appeared to be beneficial. The conditions of the two tests were not completely identical, and the differences between them do not appear in the table.

In 1956 the animals were a Duroc-Landrace cross, a strain that is quite docile. In this test they were apparently content to remain on the slab, spending very little time at the feed trough. The control pigs, which were forced by discomfort to make regular trips to the wallow were observed to spend more time eating than the slab pigs. However, they did not make as good use of this feed, as is shown by the comparative efficiency of use of feed: 360 lbs per 100 lbs of gain, as compared with 349 lbs for the slab pigs. Greater comfort of the slab pigs was also indicated by respiration rates and body temperatures; at times when the air temperature was 103 to 107 F, respiration rates for the slab pigs averaged 44 per minute and for the wallow pigs, 110. Body temperatures were 103.3 and 104.0 F, respectively, both higher than normal.

In 1957 the temperature of the slab was raised to 70 F, in an attempt to balance the comfort of sleeping with eating and thus increase feed intake. The Yorkshire hogs used in the test were considerably more restless than the Duroc-Landrace cross of the previous year, and this may have been partly responsible for increased animal movement and more frequent stops at the self-feeder.

The rate of heat loss to the slab by the pigs has measured both years by small heat-flow meters, the electromotive force being recorded on a strip chart potentiometer. Within the range of conditions of these tests little relation was found between heat flow to slab and outside air temperature.[10] At air temperatures over 100 F, rates of heat flow to the 65-degree slab average 47 hours per square foot, while the rate of flow to the air and surroundings at the same temperatures ranged from −5 to +5 Btu (British thermal units) per hour per square foot. A 150-lb hog with a total surface area of 15 square feet, of which 20% or 3 square feet was in contact with the concrete would thus be able to lose 141 Btu per hour to the slab. It has been shown by Bond et al[9] that when ambient temperature is 100 F, the total heat loss from a hog this weight is approximately 500 Btu per hour, 90% latent heat. The loss to the slab thus represents 28% of the total, a decided easing of the situation for the hog. This is also indicated by the decided lowering of the respiration rate.

Power Consumption

The amount of electric power used for the various cooling tests was measured and is summarized in Table V. The desert cooler and equipment for cooling drinking water were used with the maximum number of animals they could accommodate. The fans, however, might have cooled three times as many animals as were available for the experiments, and the cold slab twice as many hogs. This would, of course, lower considerably the power cost per animals.

Cost of equipment—initial cost, interest, depreciation, and repair—cannot

be derived from the test equipment cost, especially in the case of the refrigeration machinery for the cold drinking water and cold slab, as these were not used to capacity in all the tests. It is to be hoped, further more, that industry can develop equipment more efficient for these uses.

References

1. Map published by Frigidaire Division, General Motors Corporation, Dayton, Ohio.

2. Artificial Shades for Livestock in Hot Climates, C. F. Kelly, N. R. Ittner. *Agricultural Engineering*, St. Joseph, Mich., vol. 29, no. 6, 1948, pp. 239–42.

3. Thermal Design of Livestock Shades, C. F. Kelly, T. E. Bond, N. R. Ittner. *Ibid.*, vol. 31, no. 12, 1950, pp. 601–06.

4. Cattle Shades, N. R. Ittner, C. F. Kelly. *Journal of Animal Science*, Purdue University, Lafayette, Ind., vol. 10, no. 1, 1951, pp. 184–94.

5. Sky Temperatures in the Imperial Valley of California, C. F. Kelly, T. E. Bond, N. R. Ittner. *Transactions*, American Geophysical Union, Washington, D.C., vol. 38, no. 3, 1957, pp. 308–13.

6. "Cold Spots" in the Sky May Help Cool Livestock, C. F. Kelly, T. E. Bond, N. R. Ittner. *Agricultural Engineering*, vol. 38, no. 10, 1957, pp. 726, 729, 743.

7. Increasing Summer Gains of Livestock, N. R. Ittner, T. E. Bond, C. F. Kelly. *Journal of Animal Science*, vol. 13, no. 4, 1954, pp. 867–77.

8. Wind Machines, F. A. Brooks, D. G. Rhoades, A. S. Leonard, *California Agriculture*, Berkeley, Calif., vol. 6, no. 8, 1953, pp. 7–8.

9. Heat and Moisture Loss from Swine, T. E. Bond, C. F. Kelly, Hubert Heitman, Jr. *Agricultural Engineering*, vol. 33, no. 3, 1952, pp. 148–52.

10. The Operating Data Book—A Meter for the Evaluation of Heat Flow, J. F. Gier, R. V. Dunkle. National Association of Practical Refrigerating Engineers, Menasha, Wis., eighth edition, 1948, pp. 3–7.

A Phase-Plane Approach to Relay Sampled-Data Feedback Systems

FRANCIS J. MULLIN
ASSOCIATE MEMBER AIEE

ELIAHU I. JURY
MEMBER AIEE

Synopsis: The phase plane is used to solve second-order sampled-data feedback systems which contain a relay and zero order hold circuit. The relay quantizes the signal applied to the linear part of the system and each quantization level gives rise to a corresponding set of trajectories in the phase plane. Between sampling instants the output moves along one of these trajectories; however, the error signal which is applied to the relay depends only on the output and input at the end of the sampling interval, i.e., at the sampling instants. The output at the sampling instants is determined in the phase plane with the aid of two difference equations which are obtained from the time solution of the linear part of the system. This method of solution is not restricted to the study of such systems for step function inputs but may be used equally well to study the effects of sine wave inputs on the system as is illustrated in the examples of the paper.

SAMPLED-DATA systems which contain a relay and zero order hold circuit have been studied previously.[1–5] De Jager and Bailey[1] have indicated the possibility of using a sampler and relay to obtain a signal which is quantized in both time and amplitude for use in a pulse-code modulation system. They have discussed such systems in very general terms and did not attempt to analyze them. Altar and Helstrom[2] used the phase plane to find the output of such a system at the sampling instants. They restricted themselves to a second-order system whose characteristic equation possessed oscillatory roots. Later, Kalman[3] used the "normal co-ordinates" or partial fraction expansion to solve these systems and in doing so removed the restriction of oscillatory roots, but he was also concerned with a solution only at the sampling instants. Chow[4] and Russell,[5] on the other hand, used the describing function and obtained the approximate continuous output for sinusoidal inputs; they necessarily restricted themselves to sinusoidal inputs but were not limited to second-order systems.

In this paper second-order relay sampled-data systems will be analyzed using the conventional phase-plane techniques and by doing so the exact continuous response of such systems for step functions and sine wave inputs obtained. This approach is different from that of Altar and Helstrom and Kalman. They both used a change of variable to describe the system and were primarily concerned with the effect of initial conditions on the stability of the system. Here the primary concern is with the responses of these systems to specific forcing functions and the effects of initial conditions on the various responses will be discussed in general terms.

Paper 58-1002, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted March 13, 1958; made available for printing June 10, 1958.

Francis J. Mullin, formerly with the University of California, Berkeley, is with the California Institute of Technology, Pasadena, Calif., and Eliahu I. Jury is with the University of California, Berkeley, Calif.

This research was supported by the U. S. Air Force through the Air Force Office of Scientific Research, Air Research and Development Command, under Contract no. *AF 18(600)-1521.*

Phase-Plane Solution

Consider the system shown in Fig. 1. Let the relationship between H_n and E_n be given by

$$\begin{aligned} H_n &= 1.0 \text{ if } E_n \geq E' \\ H_n &= 0.0 \text{ if } |E_n| < E' \\ H_n &= -1.0 \text{ if } E_n \leq -E' \end{aligned} \tag{1}$$

If $E'=0$, then the relay is said to be ideal. The output of the relay, which is the input to the linear part of the system, is constant over any sampling interval and has one of the three values given in equation 1. Since $G(s)$ is assumed to be of second order, one can draw a set of trajectories relating $c(t)$ and $\dot{c}(t) \triangleq d/dt\,[c(t)]$ for each of the three values of H_n. Then, if H_n is, say -1.0, during the nth sampling interval, the operating point of the system will move along the -1.0 trajectory which passes through the point in the c, $\dot{c}$ plane which corresponds to the values of $c(t)$ and $\dot{c}(t)$ at the beginning of the sampling interval. Thus, for this sampling interval the trajectory originates at the point C_n, $\dot{C}_n$ and will terminate at the point C_{n+1}, $\dot{C}_{n+1}$ which will lie on the same trajectory if $G(s)$ is of the form: $K/(s^2+as+b)$. For notational convenience the following definitions are used:

$$C_n = c(t)\big|_{t=nT} \qquad e(t) = r(t) - c(t)$$

$$\dot{C}_n = \frac{d}{dt}[c(t)]\Big|_{t=nT} \qquad E_n = r(t)\big|_{t=nT} - c(t)\big|_{t=nT}$$

$$E_n = R_n - C_n \tag{2}$$

T is the sampling period of the system.

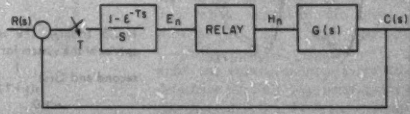


Fig. 1. A relay sampled feedback system

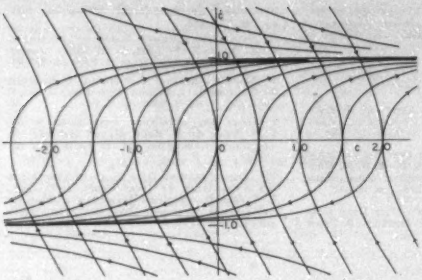


Fig. 2. Phase-plane trajectories for $H_n = \pm 1.0$ with $G(s) = \frac{1}{s(s+1)}$. The $H_n = +1.0$ trajectories approach $C = \infty$, $\dot{C} = +1.0$; the $H_n = -1.0$ trajectories approach $C = -\infty$, $\dot{C} = -1.0$

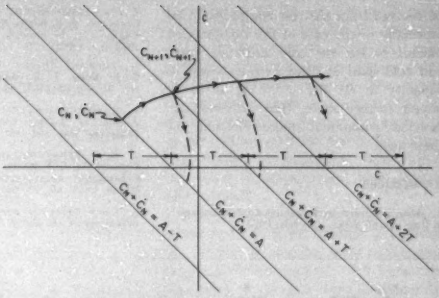


Fig. 3. Determination of C_{n+1}, $\dot{C}_{n+1}$ using the phase-plane trajectory and equation 17

Since the trajectory for any sampling interval starts at C_n, $\dot{C}_n$ and ends at C_{n+1}, $\dot{C}_{n+1}$, one must now determine exactly where the point C_{n+1}, $\dot{C}_{n+1}$ lies on the trajectory, given C_n, $\dot{C}_n$. These two points are separated in time by one sampling interval, i.e., by T seconds, but because time is not readily obtainable from a phase-plane plot, an auxiliary relationship must be found. The required information can be obtained by solving the differential equation relating $c(t)$ to H_n. The solution of this equation is of the form:

$$c(t) = H_n f_1(t) + C_n f_2(t) + \dot{C}_n f_3(t); \quad nT \le t \le (n+1)T \quad (3)$$

Also,

$$\dot{c}(t) = H_n \dot{f}_1(t) + C_n \dot{f}_2(t) + \dot{C}_n \dot{f}_3(t); \quad nT \le t \le (n+1)T \quad (4)$$

For typical forms of $f_i(t)$ see equations 13 and 21; it should also be noted that $\dot{f}_i(t) = d/dt\,[f_i(t)]$. If equations 3 and 4 are evaluated for $t = (n+1)T$, then there results two first-order difference equations which describe the system at the sampling instants. They are of the form:

$$C_{n+1} = K_1 H_n + K_2 C_n + K_3 \dot{C}_n \quad (5)$$

$$\dot{C}_{n+1} = K_4 H_n + K_5 C_n + K_6 \dot{C}_n \quad (6)$$

If equations 5 and 6 are added, the resulting equation becomes:

$$C_{n+1} + \dot{C}_{n+1} = K_a H_n + K_b C_n + K_c \dot{C}_n = K_n \quad (7)$$

$K_1 + K_4 = K_a$, etc., and obviously K_n depends upon H_n, C_n, and $\dot{C}_n$. Once K_n is determined, one simply plots equation 7 in the phase plane and the intersection of this equation with the trajectory along which the operating point of the system is moving determines the values of C_{n+1} and $\dot{C}_{n+1}$. With C_{n+1} determined E_{n+1} can be calculated and the process is then repeated. Thus, this method provides a convenient manner for determining the behavior of the system and its use will be illustrated in the following two examples.

EXAMPLE 1

In Fig. 1 let $G(s) = 1/s(s+1)$. (This is the same system investigated by both Chow[4] and Kalman.[4]) For this choice of $G(s)$ the relationship between $c(t)$ and H_n is given by

$$\ddot{c}(t) + \dot{c}(t) = H_n; \quad nT \le t \le (n+1)T \quad (8)$$

Let $v = \dot{c}$ in equation 8 and divide by v. This gives:

$$\frac{\dot{v}}{\dot{c}} = \frac{dv}{dc} = \frac{H_n - v}{v} \quad (9)$$

Separating terms in equation 9 shows:

$$\int_{C_n}^{c} dc = \int_{V_n}^{v} \frac{v\,dv}{H_n - v} \quad (10)$$

Integrating this equation, it is found that:

$$c - C_n = -\left[(v - V_n + H_n \ln\left(\frac{H_n - v}{H_n - V_n}\right)\right] \quad (11)$$

Equation 11 gives the desired relation-

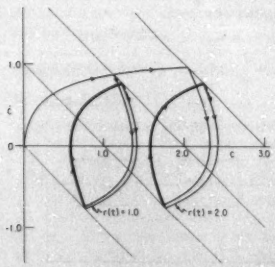
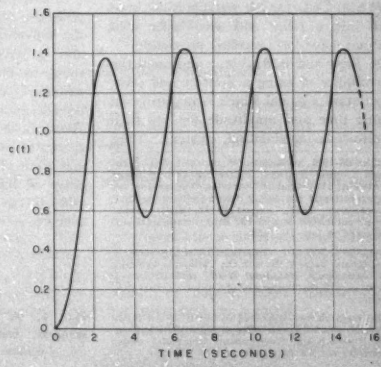


Fig. 4 (left). Response of the system with an ideal relay, T = 1.0 second and $G(s) = \frac{1}{s(s+1)}$ to r(t) = 1.0 and 2.0

Fig. 5 (right). The time response of the system for T = 1.0 second and $G(s)\frac{1}{s(s+1)}$ to r(t) = 1.0



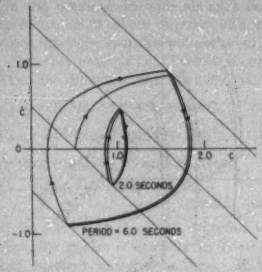


Fig. 6. Limit cycles with periods of 2.0 and 6.0 seconds for the same system with $r(t)=1.0$

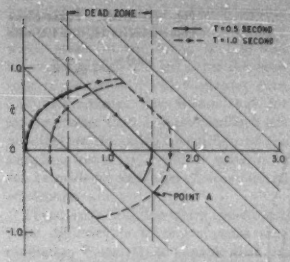


Fig. 7. Response of the system when the relay has a dead zone and $r(t)=1.0$

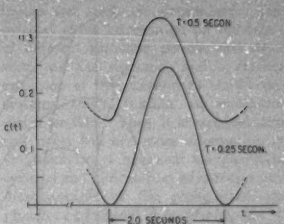


Fig. 9. Steady-state response of the system to $r(t)=1.0 \sin \pi t$

ship between c and $v(=\dot{c})$ necessary to plot the phase-plane trajectories. For the moment assume that the relay in Fig. 1 is an ideal relay. ($E'=0$ in equation 1.) Then there are two sets of trajectories in the phase plane, one set for $H_n=1.0$ and the second set for $H_n=-1.0$; these trajectories are shown in Fig. 2.

Next, the time solution of equation 8 will be obtained. First, the time scale is shifted so that $t=nT$ corresponds to $t'=0$ where $t'=t-nT$ and $0\leq t'\leq T$. Then, the Laplace transform of equation 8 is

$$\mathcal{L}[c(t')] = C(s) = \frac{H_n}{s^2(s+1)} + \frac{C_n}{s} + \frac{\dot{C}_n}{s(s+1)} \quad (12)$$

The inverse of equation 12 is

$$c(t) = H_n(t'-1+e^{-t'}) + C_n + \dot{C}_n(1-e^{-t'}); \quad nT\leq t\leq(n+1)T \quad (13)$$

and

$0\leq t'\leq T$

where

$t'=t-nT$

Differentiating equation 13 gives:

$$\dot{c}(t) = H_n(1-e^{-t'}) + \dot{C}_n(e^{-t'}) \quad (14)$$

Lettering $t'=T$ in the equations 13 and 14 gives the two first-order difference equations which describe the system at the sampling instants. They are

$$C_{n+1} = H_n(T-1+e^{-T}) + C_n + \dot{C}_n(1-e^{-T}) \quad (15)$$

$$\dot{C}_{n+1} = H_n(1-e^{-T}) + \dot{C}_n(e^{-T}) \quad (16)$$

Adding equations 15 and 16 gives:

$$C_{n+1} + \dot{C}_{n+1} = TH_n + C_n + \dot{C}_n \quad (17)$$

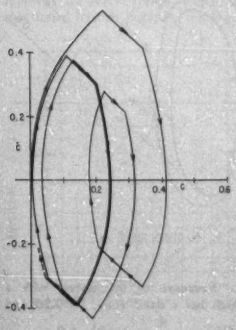
Refer now to Fig. 3. Assume that C_n and $\dot{C}_n$ are such that their sum is A and hence the point C_n, $\dot{C}_n$ lies along the line $C_n+\dot{C}_n=A$. Then, if H_n is $+1.0$, the operating point will move along one of the $+1.0$ trajectories shown in Fig. 2 until the trajectory intersects the line $C_n+\dot{C}_n=A+T$ and the intersection of the trajectory with this line determines the $\dot{C}_{n+1}$, C_{n+1} point. If H_{n+1} is $+1.0$, the operating point continues along the same trajectory until it intersects the line $C_n+\dot{C}_n=A+T+T=A+2T$; if H_{n+1} is -1.0, the operating point moves along the -1.0 trajectory which passes through the C_{n+1}, $\dot{C}_{n+1}$ point until the trajectory intersects the line $C_n+\dot{C}_n=A+T-T=A$. Thus, at the sampling instants the system operating point is always along one of the lines shown in Fig. 3 if it initially starts on one of these lines; a similar statement holds for any other choice of initial conditions.

The phase-plane response of this system, assuming an ideal relay, to step functions of two magnitudes is shown in Fig. 4 with $T=1.0$ second, and the time response of the system for $r(t)=1.0$ is shown in Fig. 5. From these two responses it is evident that the system is unstable and will exhibit a limit cycle for step inputs of all magnitudes. However, the period of the oscillations definitely depends on the choice of initial conditions. The period of the limit cycles shown in Fig. 4 is 4 seconds while Fig. 6 shows choices of initial conditions which lead to oscillations with periods of 2 seconds and 6 seconds when $r(t)=1.0$. These results have been reported previously by Kalman and for more information on the effect of initial conditions on the period of oscillations, the reader is referred to reference 3. However, the phase plane clearly shows these limit cycles and the period of oscillation for any choice of initial conditions can be determined without much difficulty.

If the relay is permitted to have a dead zone, a stabilizing effect can be achieved. Let E' in equation 1 be 0.5; then:

$$\begin{aligned} &H_n=1.0 \text{ if } R_n-C_n\geq 0.5 \\ &H_n=0 \text{ if } |R_n-C_n|<0.5 \\ &H_n=-1.0 \text{ if } R_n-C_n\leq -0.5 \end{aligned} \quad (18)$$

The trajectories for $H_n=0$ are straight lines whose equations are given by equation 11 with $H_n=0$. Also, since $C_{n+1}+\dot{C}_{n+1}=C_n+\dot{C}_n$ when $H_n=0$, one must use either equation 15 or 16 to determine the C_{n+1}, $\dot{C}_{n+1}$ point, given the C_n, $\dot{C}_n$ point when the system operating



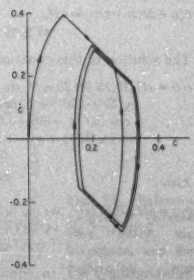


Fig. 8. Response of the system when the relay has a dead zone and $r(t)=1.0 \times \sin \pi t$

A—T=0.25 second
B—T=0.50 second

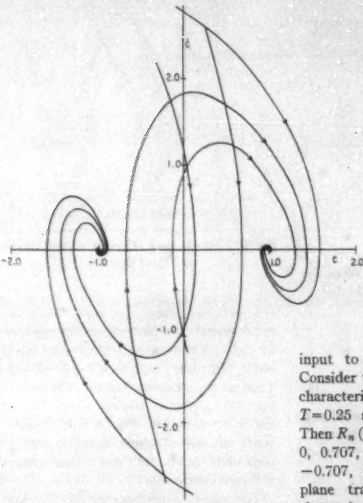


Fig. 10 (left). Phase-plane trajectories for $H_n = \pm 1.0$ with $G(s) = \frac{4}{s^2+2s+4}$

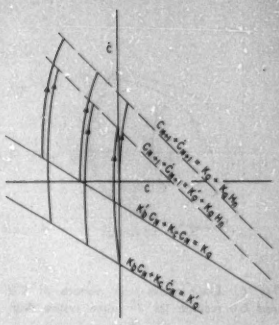


Fig. 11 (right). A suggested scheme for determining K_n

point is moving along one of the $H_n=0$ trajectories. The response of the system to a unit step function is shown in Fig. 7 and it is seen that the system is stable for $T=0.5$ second and unstable for $T=1.0$ second.

There are several interesting points which should be observed in Fig. 7. For zero initial conditions (and this will be assumed to be the case in all of what follows unless it is specifically stated otherwise) the only stable steady-state value of $c(t)$ for $T=0.5$ or 1.0 second is 1.0. This is due to the fact that the system operating point at the sampling instants must lie on one of the lines $C_n+\hat{C}_n=0$, ± 0.5, ± 1.0, ± 1.5, etc. (depending upon T); the only one of these lines which intersects the dead zone is the line $C_n+\hat{C}_n=1.0$ and the point $C_n=1.0, \hat{C}_n=0$ is the equilibrium point on this line. Also, if the system is to be stable for $r(t)=1.0$ and $T=1.0$ seconds, then Fig. 7 indicates that point "A" must lie in the dead zone and so the width of the dead zone required for the system to be stable to this input can be determined from Fig. 7. The width of the dead zone required for stability for step inputs of other magnitudes can be determined in a similar fashion and hence it is possible to find the dead zone required to make the system stable for step inputs of any magnitude and for any valve of the sampling period.

The same method can be used if the input to the system is a sine wave. Consider the same system with the relay characteristics given by equation 18, $T=0.25$ second and $r(t)=1.0 \sin \pi t$. Then R_n (see equation 2) has the sequence 0, 0.707, 1.0 0.707, 0, -0.707, -1.0, -0.707, 0, 0.707,. . . . and the phase-plane trajectories for this input are shown in Fig. 8(A). If $T=0.5$ second, R_n has the sequence 0, 1.0, 0, -1.0, 0, 1.0,. . . . The phase-plane solution for the two values of T is shown in Fig. 8(B) and the continuous time responses for the two values of T are shown in Fig. 9. It is noted that there is a d-c component in the output which might not be noticed if the describing function were used to analyze the system.

EXAMPLE 2

Now consider the system shown in Fig. 1 with $G(s)=4/s^2+2s+4$. Following the procedure used in Example 1 it is found that:

$$\frac{\dot{v}}{\dot{c}}=\frac{dv}{dc}=\frac{4H_n-4c-2v}{v} \qquad (19)$$

The trajectories for $H_n=+1.0$ and -1.0 are shown in Fig. 10. The differential equation relating $c(t)$ to H_n is

$$\ddot{c}(t)+2\dot{c}(t)+4c(t)=4H_n; \quad nT\leq t\leq(n+1)T \qquad (20)$$

The solution of this equation is

$$c(t)=4H_n[0.25+0.289e^{-t'}\sin(\sqrt{3}t'-120)]+1.155C_ne^{-t'}\cos(\sqrt{3}t'-30)+0.577\dot{C}_ne^{-t'}\sin(\sqrt{3}t'); \quad t'=t-nT \text{ and } 0\leq t'\leq T \qquad (21)$$

Also

$$\dot{c}(t)=4H_n[0.577e^{-t'}\sin(\sqrt{3}t')]-2.31C_ne^{-t'}\sin(\sqrt{3}t')+1.155\dot{C}_ne^{-t'}\cos(\sqrt{3}t'+30) \qquad (22)$$

Substituting $t'=T$ in equations 21 and 22 gives the two difference equations which describe this system; they have the same form as equations 5 and 6 and their sum is of the same form as equation 7. To simplify the calculation of K_n in equation 7 it is suggested that the following procedure be used. The sum of equations 5 and 6 can be written as:

$$C_{n+1}+\hat{C}_{n+1}-K_aH_n=K_bC_n+K_c\hat{C}_n=K_n \qquad (7A)$$

If $K_bC_n+K_c\hat{C}_n=K_n$ is plotted in the phase plane for various values of K_n, then the value of K_n for any point in the phase plane can be accurately estimated by inspection. With K_n determined it is again a simple matter to find C_{n+1} and $\hat{C}_{n+1}$ by plotting the auxiliary relationship $C_{n+1}+\hat{C}_{n+1}=K_n+K_aH_n$ for the same values of K_n. Such a procedure is illustrated in Fig. 11 for two values of K_n.

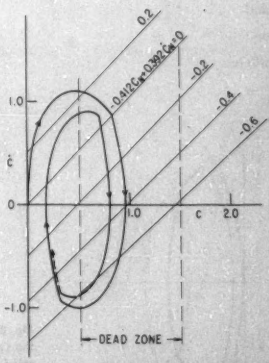


Fig. 12. Response of the system with a relay which has a dead zone, $T=0.50$ and $G(s)=\frac{4}{s^2+2s+4}$ to $r(t)=1.0$

(In the previous example all such lines made an angle of 45 degrees with the c-axis but in general this will not occur.)

For this example let $T=0.50$ second. Then the difference equations become:

$$C_{n+1}=0.345H_n+0.654C_n+0.266\dot{C}_n \quad (23)$$

$$\dot{C}_{n+1}=1.066H_n-1.066C_n+0.126\dot{C}_n \quad (24)$$

Adding equations 23 and 24 gives:

$$C_{n+1}+\dot{C}_{n+1}=1.411H_n-0.412C_n+0.392\dot{C}_n \quad (25)$$

If the relay is ideal, a little reflection will show the system is unstable if $r(t)=0$ for any nonzero initial conditions. If the relay has a dead zone, a different situation may occur. First of all, there is an additional set of trajectories corresponding to $H_n=0$ which spiral in toward the origin in a manner similar to those shown in Fig. 10. To be specific, let the relay characteristics again be given by equation 18 and let $r(t)$ be a unit step function. To solve this problem the equation $-0.412C_n+0.392C_n=K_n$ is plotted in the phase plane for various values of K_n so that the right side of equation 25 can be evaluated with a minimum of difficulty. Once this is done, the solution may be carried out in the manner previously described. The phase-plane solution of this problem is shown in Fig. 12 and it is evident that the system is unstable for this input. However, if the magnitude of the step function input is greater than 1.70, the system will not oscillate but will have a steady-state value of 1.0. This is because the maximum value of c on the trajectory which starts at the origin is 1.20 and hence the system will not enter the dead zone of the relay if $r(t)$ is greater than 1.70. Essentially, this means that for inputs greater than 1.70 the feedback has no effect on the output regardless of the sampling frequency. Similar statements can be made for any other choice of initial conditions.

There is one situation which could cause some difficulty in solving systems with oscillatory roots. It is possible that the line $C_{n+1}+\dot{C}_{n+1}=K_n$ might intersect the trajectory at more than one point. If this occurs, the selection of the proper point of intersection can usually be made from a knowledge of how much the system can change during one sampling interval. If there is a doubt, then one must compute the value of C_{n+1} or $\dot{C}_{n+1}$ from the linear difference equation which describes the system.

Conclusions

This paper extends the phase-plane approach to the analysis of relay sampled-data control systems. For the second-order systems discussed this approach readily yields the stability conditions and limit cycles of the systems as well as the response at all times. Step function and sinusoidal inputs are discussed in the examples; however, extension to other inputs can easily be determined. The method yields information on the effect of different sampling periods and dead zones on the response, and a further extension of this method to third-order systems which contain an integrator seems feasible. Thus, it is expected that this paper will add more information to the behavior of nonlinear sampled-data systems and indicate the advantages and disadvantages of the phase-plane approach as compared to other methods.[4-7]

References

1. *Discussion* by **F. De Jager** and **A. E. Bailey** of Characteristics of Sampling Servo Systems, Smith, Lawden, Bailey, "Automatic and Manual Control" (book), A. Tustin, editor. Butterworths Scientific Publications, London, England, 1952, pp. 406-07.

2. Phase Plane Representation of Sampling Servomechanisms, W. Altar, C. W. Holstrom. *Research Report R-94410-14-B*, Westinghouse Research Laboratories, Pittsburgh, Pa., Sept. 1953.

3. Some Aspects of Nonlinear Sampled-Data Systems, **R. E. Kalman**. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1956, pp. 276-313.

4. Contactor Servomechanisms Employing Sampled Data, **C. K. Chow**. *AIEE Transactions*, vol. 73, pt. II, Mar. 1954, pp. 51-62.

5. *Discussion* by **Frederick A. Russell** of reference 4, pp. 63-64.

6. Nonlinear Feedback Systems with Sampling, **E. Mishkin, M. S. Goldstein, J. G. Truxal**. *Research Report R-450-55, PIB-340*, Polytechnic Institute of Brooklyn, Aug. 1956.

7. Analysis and Design of Nonlinear Sampled-Data Control Systems, **Robert F. Nease**. *WADC Technical Note 57-162*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., June 1957.

Discussion

Keisuke Izawa (Purdue University, Lafayette, Ind.): The method developed by the authors for the analysis of relay sampled-data control systems is valuable and simple, their results indicating that the phase-plane approach is effective to trace the time behavior when sampling action is followed by relay action, which was also mentioned by this discusser in his paper.[1]

For the relay sampled-data control systems, the cyclic response (including the almost periodic response as well as the periodic one) must be considered permissible, if the maximum error deviation is within tolerable limits. Such effort to limit the error, regardless if it is steady still or cyclic, within the allowable range for any possible initial condition should be taken in the design of the control system.

In the control system with dead zone (Example 1 of the paper), the trajectory associated with the dead zone may vanish if

$$a<\frac{\lambda}{2}, \qquad \lambda \leq T \quad (26)$$

where a is the half-width of the dead zone and T is the sampling time. For this case the over-all trajectory will look like the one in Fig. 6 or the one of the case without dead zone. The approach taken in the paper[2] by Weaver and the discusser yields the following formulas for cyclic response

$$a=e^{\lambda}\tanh\left(\frac{P_h}{2}\right)-\frac{P_h}{2}-e^{\lambda}+1+\lambda \quad (27)$$

$$x_m=\ln\left\{\cosh\left(\frac{P_h}{2}\right)\right\} \quad (28)$$

where P_h is the half-period of the cyclic response, x_m is the amplitude of the cyclic response, and a and λ have to fulfill condition 26.

Equation 28 can also hold for the case without dead zone. For the first case (without dead zone) of Example 1, the fact that only three pure periodic modes corresponding $P_h=1$, 2, and 3 exist is easily verified.

Fig. 13 shows the discusser's calculation on the authors' first problem. Curve I indicates the relation between variable delay λ due to the sampling and the half-period

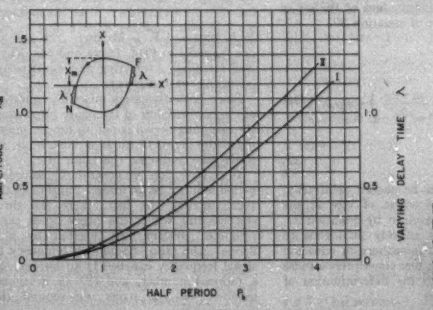


Fig. 13. Relations between variable time delay, the amplitude, and the half-period from Example 1

P_h, which is the time duration between two switching points N and F, given by

$$e^{\lambda} \tanh\left(\frac{P_h}{2}\right) - \frac{P_h}{2} - e^{\lambda} + 1 + \lambda = 0 \qquad (29)$$

Three cases of pure periodic responses shown in Figs. 4 and 6 of the paper can easily be verified in the following way. Since λ may vary between 0 and 1,[1,2] the possible P_h will be any integral multiple of the sampling time between 0 and 3.750 (these two numbers correspond to $\lambda = 0$ and 1 respectively). Therefore, the half-periods of the possible pure periodic motion will be 1, 2, and 3, and the corresponding amplitudes are 0.120, 0.434, and 0.851. Furthermore, if the sampling time is changed from 1.0 to 0.4 in the same case, λ may vary from 0 to 0.4 and the possible half-periods are obtained from curve I of Fig. 13 as 0.4, 0.8, 0.16, and 2.0. The corresponding amplitudes are 0.0198, 0.0779, 0.290, and 0.434. With curve II of the figure, the amplitude x_m, and the half-period P_h are related.

The authors discuss the pure periodic responses of the relay sampled-data control system. These responses, in turn, require the definite regions of initial point on the phase plane. The sinusoidal input whose period is integral multiple of the sampling period is also considered in their study. It would be appreciated if the author would explain whether or not there occurs any significant fact or change in their approach for such general case of not purely periodic response and/or when the period of the sinusoidal input is not the integral multiple of the sampling period; these conditions seem to take place with much probability.

REFERENCES

1. ON-OFF CONTROL WITH PERIODIC SENSING DEVICE, **K. Izawa.** *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 80, no. 7, 1958, pp. 1459–64.

2. RELAY-TYPE FEEDBACK CONTROL SYSTEMS WITH DEAD TIME AND SAMPLING, **Keisuke Izawa, L. E. Weaver.** *AIEE Transactions*, vol. 78, pt. II, 1959 (*Paper no. 58-1026*).

Fred Kurzweil, Jr. (Stanford University, Stanford, Calif.): The authors are to be commended for a very able presentation of a method for identifying sampling times in the phase plane. A similar study is presently being conducted at Stanford University, to which the method finds direct application.

Two comments might be added. The first is a comment regarding the use of sinusoidal analysis techniques in the phase plane as was done in Example 1, Figs. 8(A) and 8(B). For this example, a fortuitous choice of system parameters and system input led to a response which reached its steady state after a few (comparatively speaking) sampling periods. (Note that the period of the applied sinusoid is a multiple of the sampling period.) Thus analysis in the phase plane for this case proved to be simple and straightforward. In general, however, these restrictions do not hold and sinusoidal analysis in the phase plane becomes an extremely tedious matter, requiring the plotting of many points for each value of input amplitude and input frequency before the behavior of the system can be established. Moreover, this type of analysis does not remotely lead to any closed expression relating the output of the system to its input. Therefore, in general, the phase plane does not lend itself readily to sinusoidal analysis of nonlinear systems, and although the method proposed in the paper reduces the amount of labor required to compute individual points on a trajectory, the same is true here.

The second comment has to do with an extension of the method proposed by the authors.

Consider a sampled system described by the following set of difference equations:

$$C_{n+1} = a_{11}C_n + a_{12}\dot{C}_n + A_1$$
$$\dot{C}_{n+1} = a_{21}C_n + a_{22}\dot{C}_n + A_2 \qquad (30)$$

These equations are of the same form as equations 5 and 6 used by the authors to describe the system of Fig. 1 for $G(s)$ of second order. A_1 and A_2, which describe the forcing function of the system at sampling instants, can be functions of time; however, as the authors demonstrate, in the practical case these will be constant over several sampling intervals, i.e., the system will be operating along one trajectory. Thus, we consider the case for A_1 and A_2 constant. The equations are multiplied by arbitrary constants, k_1 and k_2, and then added:

$$k_1C_{n+1} + k_2\dot{C}_{n+1} = (k_1a_{11} + k_2a_{21})C_n + (k_1a_{12} + k_2a_{22})\dot{C}_n + k_1A_1 + k_2A_2 \qquad (31)$$

If the condition is imposed that:

$$k_1a_{11} + k_2a_{21} = kk_1$$
$$k_1a_{12} + k_2a_{22} = kk_2 \qquad (32)$$

where k is a constant to be determined, equation 31 becomes:

$$k_1C_{n+1} + k_2\dot{C}_{n+1} = k(k_1C_n + k_2C_n) + k_1A_1 + k_2A_2 \qquad (33)$$

or

$$S_{n+1} = kS_n + A \qquad (34)$$

This difference equation expressed in terms of initial values and for A constant is

$$S_n = (k)^n S_0 + A\sum_{r=0}^{n-1} k^r \qquad n = 1, 2, \ldots . \qquad (35)$$

Thus, $k_1C_n + k_2\dot{C}_n$ become a series of parallel lines with spacing determined by equations 34 or 35.

The solution of equation 32 reveals that k is identical to the eigenvalues of the set of difference equations of equation 30.

$$\lambda_{1,2} = k = \frac{a_{11} + a_{22}}{2} \pm \sqrt{\left(\frac{a_{11} - a_{22}}{2}\right)^2 + a_{12}a_{21}} \qquad (36)$$

$$(a_{11} - k)k_1 = -k_2a_{21}$$
$$(a_{22} - k)k_2 = -k_1a_{12} \qquad (37)$$

Either k_1 or k_2 can be chosen arbitrarily, but for a specified value of one k, the remaining k is determined by equation 37.

The foregoing is entirely analogous to that suggested by the authors in equation 17 and Fig. 3 and in fact constitutes a more general approach to the determination of the $k_1C_n + k_2\dot{C}_n$ lines.

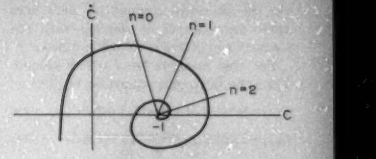


Fig. 14. Intersection of continuous trajectory with rays

Since there are two values of k (from equation 36), there will be two sets of parallel lines. The points of intersection of the two sets of parallel lines correspond to the solutions of equation 30 for given sampling times, n. The fact that two sets of parallel lines occur should occasion no difficulty; each set of parallel lines corresponds (by proper transformation) to the eigenvector associated with k. The eigenvectors are displaced according to the law expressed in equation 35.

If the eigenvalues are complex, then the foregoing method no longer is applicable, and to identify time in the phase plane, a scheme such as that illustrated in Fig. 11 is required. Still another scheme might be mentioned. With reference to Fig. 10 as an example, it is observed that the continuous trajectories spiral around either the $+1$ or the -1 point. If the increments of time are to be measured on a given trajectory, this is accomplished by drawing rays from the $+1$ or the -1 point and noting the intersection of the ray with the continuous trajectory, as shown in Fig. 14. Time is identified by measuring the angle between the c-axis and the ray and expressing this angle as a function of time. By an appropriate transformation from a normal coordinate phase space, this is found to be

$$\phi_n = \tan^{-1} \times \left[\begin{array}{c} a_{11} - \beta\,(C_0(a_{11} - \beta \tan nx) - \dot{C}_0 a_{12}) \\ a_{12} - a_{12}(C_0(\beta + a_{11} \tan nx - \tan nx) + \\ \dot{C}_1 a_{12} \tan nx) \end{array}\right]$$

$\phi \triangleq$ angle between C-axis and ray
$n \triangleq$ sampling instant
$k = \alpha \pm \beta j \triangleq$ eigenvalues of the system
$a_{ij} \triangleq$ parameters of equation 30
$x = \tan^{-1} \beta/\alpha$
$C_0 =$ initial value of displacement
$\dot{C}_0 =$ initial value of velocity

E. I. Jury and **F. J. Mullin:** The authors are thankful to Prof. Izawa and Mr. Kurzweil for their interesting and constructive comments on our paper.

Since both discussers raised the question of the applicability of the phase plane when the sampling period is not an integer multiple of the period of the input sine wave, we would like to discuss this point first.

It is true that in general the sampling period will not be an integer multiple of the signal period. In fact, there are three relationships which can occur when sampled-data systems are excited by sine waves: the ratio of sampling frequency to signal frequency can be (1) an integer, (2) a rational number, and (3) an irrational number. In the paper we showed the

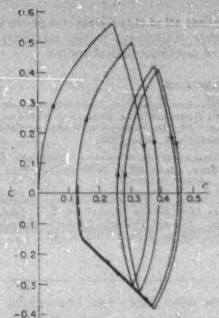


Fig. 15. The phase-plane response to $r(t)=2.0 \sin \pi t$ with $T=0.8$ second and $G(s)=\frac{1}{s(s+1)}$

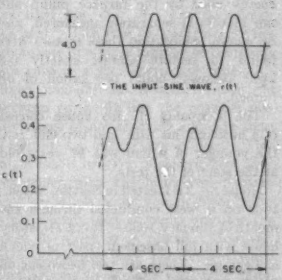


Fig. 16. The steady-state time response with $r(t)=2.0 \sin \pi t$

applicability of the phase plane when this frequency ratio was an integer and further comments on this case are not warranted here. We would like to point out, however, that our method of analysis is still applicable in the more general cases but it becomes more involved in view of the nature of the sampling conditions.[1]

To illustrate this, in Fig. 1 of the paper let the relay characteristics be given by equation 18, let $G(s)=1.0/s(s+1)$, $r(t)=2.0 \sin \pi t$ and $T=0.8$ second; this gives a frequency ratio of 5/2. The phase-plane response for these conditions is shown in Fig. 15 and in this case it is observed that the period of the output is 4.0 seconds or twice the period of the input. The steady-state time response is shown in Fig. 16. The reason for this is that the period of the sampled input is 4.0 seconds and since the system responds only to the sampled input, it seems natural to expect the period of the output to also be 4.0 seconds.

To demonstrate the response when the ratio of sampling frequency to signal frequency is an irrational number, let the relay characteristics be given by equation 18, let $G(s)=1.0/s(s+1)$ in Fig. 1 of the paper, $r(t)=1.0 \sin \sqrt{2}\pi t$ and $T=0.5$ second. This gives a frequency ratio of $2\sqrt{2}$. The initial portion of the phase plane response is shown in Fig. 17 and the corresponding time response in Fig. 18. It is obvious that the output is not a periodic function and there is no reason to suspect that it would be periodic since the sampled input is not a period function.

At this point it should be noted that the standard describing function is not applicable in these two cases and leads us to conclude that the describing function can be successfully used only when the ratio of sampling frequency to signal frequency is an integer, assuming the system is not capable of self-sustained oscillations. The phase plane can be used to determine whether or not this last condition is satisfied.

In the paper we included only the case where the sampling frequency to signal frequency ratio was an integer for two reasons. First, it is the simplest case and since we were interested in explaining the phase plane as a method of analysis, we felt it best not to complicate the issue by including the two cases covered in this closure. Second, we thought that it should be obvious that this same method can be used to find the approximate output of continuous feedback systems which contain a relay when they are subjected to sinusoidal inputs. It is generally known[2] that continuous relay feedback systems can be approximated by the sampled data system shown in Fig. 1 of the paper. In this case the sampler and hold circuit become fictitious and for such an approximation one has the choice of the sampling period at his disposal. By selecting the sampling period as a submultiple of the period of the input sine wave, the phase plane can be used to analyze second-order systems in a straightforward manner.

It is of interest to note the parallel and independent work of Prof. Izawa on this problem and his clarifications and additions are certainly welcome. Regarding his comments on the effects of initial conditions on the response we would like to say that the phase plane is best suited for the response of the system with given specific initial conditions. However, it is possible to "break up" the phase plane into regions which give rise to various periodic (or almost periodic) responses as follows. Assume for simplicity the relay in Fig. 1 is ideal. Then H_n must equal $+1.0$ or -1.0. Assume we wish to make known the regions in the phase plane which give rise to the $+1.0$, -1.0, $+1.0$, -1.0, etc., sequence for H_n. If this were the sequence for H_n, then the z-transform of H_n would be

$$Z[H_n]=H^*(z)=1-z^{-1}+z^{-2}-z^{-3}+z^{-4}-z^{-5}+H^*(z)=\frac{z^2}{z^2-1}-\frac{z}{z^2-1} \quad (38)$$

With $G(s)=1.0/s(s+1)$, the sampled output of the system can be written as:

$$C^*(z)=G^*(z)H^*(z) \quad (39)$$

where

$$G^*(z)=Z\left[\frac{1-e^{-Ts}}{s}G(s)\right]$$

and

$$G^*(z)=\frac{0.368z+0.264}{z^2-1.368z+0.368}$$

for $T=1.0$ second.

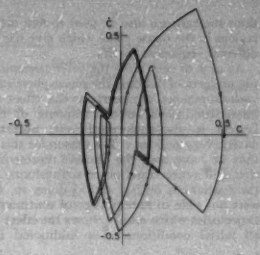


Fig. 17. The initial portion of the phase-plane response to $r(t)=1.0 \sin \sqrt{2}\pi t$ with $T=0.5$ second and $G(s)=\frac{1}{s(s+1)}$

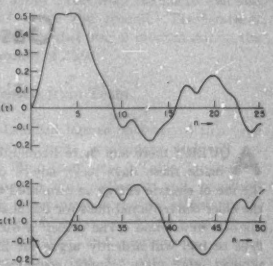


Fig. 18. The initial portion of the time response to $r(t)=1.0 \sin \sqrt{2}\pi t$

Substituting this expression into equation 39 gives:

$$C^*(z)=\frac{0.368z+0.264}{z^2-1.368z+0.368}H^*(z) \quad (40)$$

In terms of a difference equation, equation 40 can be interpreted as meaning:

$$C_{n+2}-1.368C_{n+1}+0.368C_n=0.368H_{n+1}+0.264H_n \quad (41)$$

The problem now is to find those values of C_0 and C_1 which will give rise the assumed sequence for H_n. C_1 can be easily related to C_0 (see equation 15 of the paper) so that specifying C_1 is equivalent to specifying C_0. For given values of C_0 and C_1 the solution to equation 41 is

$$C_n=0.5+1.582C_1-0.582C_0+0.038(-1)^n+(0.368)^n 1.986-C_0+C_1 \quad (42)$$

Thus we see that the two conditions which must be satisfied by C_0 and C_1 are

$$0.5+1.582C_1-0.582C_0+0.038+(0.368)^n[1.986-C_0+C_1]<0 \quad (43)$$

$$0.5+1.582C_1-0.582C_0-0.038+(0.368)^n[1.986-C_0+C_1]>0 \quad (44)$$

for all $n\geq 2$. By allowing $n\to\infty$, one finds the limiting cases which must still be checked for intermediate values of n. This

same method can also be used to find other regions in the phase plane which give rise to different sequences for H_s.

As may be seen from the brief discussion, the problem of dividing the phase plane into various regions to ascertain the effect of initial conditions on the response of the system is straightforward but, unfortunately, very tedious. The reason for this is that we have used the discrete representation (difference equation or z-transform) of the continuous output and in doing so, we were not able to obtain a set of continuous trajectories which clearly shows the effect of all initial conditions. For additional information on the subject the reader is again referred to reference 3 of the paper.

Mr. Kurzweil's comments on extending our technique to more complicated systems certainly warrants further investigation, and the forthcoming paper he mentions appears to be a timely one. We would like to add that since our paper was written, one of the authors has extended the techniques of the paper to third-order systems which contain an integrator and second-order systems which contain a pure time delay.[1]

In conclusion we want to state that we appreciate the efforts of the discussers in adding to the material of the paper and we hope that all of these ideas will prove useful to other researchers in the relatively new area of nonlinear digital control.

References

1. The Analysis and Compensation of Nonlinear Sampled-Data Feedback Systems, E. J. Mullin. *Ph. D. Thesis*, University of California, Berkeley, Calif., Sept. 1958.
2. Sampled-Data Control Systems (book), E. I. Jury. John Wiley & Sons, Inc., New York, N. Y., 1958.
3. A Note on the Operational Solution of Linear Difference Equations, E. I. Jury, E. J. Mullin. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 266, no. 3, Sept. 1958, p. 189.

Residential Heating: Electric Versus Gas or Oil

G. S. SMITH
FELLOW AIEE

A QUERY more and more frequently made these days seeks advice on the use of electric energy as compared to the older conventional fuels for the heating of a new home. The dealer in each type of fuel will ardently argue for his product and often supply convincing evidence. While there is a great deal of information available on each type as used in different houses, a meaningful comparison is difficult if not impossible because of the great number of variables involved.

About 2 years ago the author was asked to supervise and utilize the data to be taken on two residences, in each of which was installed adequate electric heating in addition to the original central furnaces. The two heating methods were to be used alternately, with careful measurements of all fuel used for each. Unfortunately, because of sickness in the family, the occupants of one house were advised by their doctor to move to a drier climate and this portion of the test was discontinued about a month after it was started. However, the results obtained during this month are included in this report for what they may be worth, and afford an excellent illustration of the results to be expected because of the variation in the occupants' living habits, differences in house construction, and so on.

Description of Test Houses

The house in which a very complete set of tests was obtained will be termed House no. 1. These tests were started in January 1956 and continued for about a month. They were again started about the first of October in 1956 and continued until the latter part of December. In 1957 they were started early in October and continued to the middle of March 1958.

House no. 1, located in Yakima, Wash., is a rectangular one-story house with all of the main floor considered in this test, except a small bathroom in one corner of the house which was originally heated electrically and not by the furnace. The electric energy used by the bathroom is metered on the house meter and not on the special test meter.

The heat portion under test covers about 1,087 square feet with a volume of about 8,654 cubic feet. Its 191 square feet of window area is largely supplied with storm windows. The construction throughout is good; it includes standard lath and plaster walls and ceilings. A 3.5-inch coat of rock wool provides heat insulation over all the ceiling area. The attic above is ventilated to the outside air. Only 25% of the total area below the house is excavated for furnace space; the remainder has adequate crawl space.

The furnace originally installed is an air-type unit, using oil, and is controlled by an electric clock thermostat. It has hot and cold air ducts to each room. The nozzle capacity was set at slightly less than one gallon per hour. The house and furnace were about 3 years old at the time these tests were started.

The metered electric heating consists of 12.35 kw of radiant-type panels with each room controlled by its individual thermostat. This load is measured by a special meter; likewise, the electric energy used by the furnace pump and fan motors is separately measured, and by careful calibration these motors were found to require almost exactly 0.8 kwhr (kilowatt-hour) per gallon of oil burned.

The occupants of this house consist of a husband, his wife, and two children, the younger of whom was 12 years old at the start of the tests.

House no. 2 is located in Seattle, Wash.; tests on it were conducted through the month of January 1956. It is a 1½-story house with full basement, part of which is a garage. There are two bedrooms on the second floor. The construction is of good standard quality, but the house has no storm windows or heat insulation.

The central hot-air heating system is a cast-iron coal furnace converted to use gas as fuel. This furnace is used to heat the first floor only with the furnace control in the living room. The two bedrooms upstairs were always electrically heated and their electric energy is measured by the house meter and not included in test meter readings.

The main floor, used in these tests, has a 906-square-foot area and a 7,094-cubic-foot volume. It is equipped with 9.25 kw of radiant-type panels with controlling thermostats in each room. The electric energy used by these panels is measured by a special meter, while the gas burned is measured by means of the regular gas meter.

The occupants of this house consist of

Paper 58-1012, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted May 9, 1958; made available for printing June 24, 1958.

G. S. Smith is with the University of Washington, Seattle, Wash.

Table I. Sample of Test Results*

Date	Type of Fuel	Furnace Motors, Kwhr	Heating, Kwhr or Gallon Oil	Unim Fuel Heat	Degree-Days	Units per Degree-Day per 1,000 Cubic Foot	Units per Degree-Day per 1,000 Square Feet	Kwhr per Degree-Day per 1,000 Cubic Foot	Kwhr per Degree-Day per 1,000 Square Feet	Average Outside Temperature
Dec. 16–23, 1957	Oil	31.6	39.5	533.5	193.9	0.318	2.53	0.930	7.40	37.3
Dec. 23–30, 1957	Electricity		510.	174.2	193.2	0.104	0.830	0.305	2.33	37.4
Dec. 30–Jan. 6, 1957–58	Oil	38.9	48.6	656.0	262.0	0.290	2.31	0.847	6.75	29.0
Jan. 6–13, 1958	Electricity		681.	232.5	233.1	0.115	0.917	0.337	2.69	31.7
Jan. 13–20, 1958	Oil	32.1	40.0	540.0	202.3	0.308	2.46	0.903	7.19	36.1
Jan. 20–27, 1958	Electricity		516.	176.3	196.7	0.104	0.828	0.303	2.42	36.9

*Unit fuel heat = 10,000 Btu
1 kwhr = 3,415 Btu
Oil = 135,000 Btu per gallon
Gallon oil used = furnace kwhr/0.8
Heated house area = 1,087 square feet
Heated house volume = 8,554 cubic feet
Installed heating = 12.35 kw

three adults and one child, who was 7 years old at the start of the tests.

Method of Conducting Tests

For the first month the tests on these two houses were conducted with oil or gas used one day and electric energy the next. This type of operation has the advantage of a large number of comparative runs in a short period of time with slight changes in weather conditions between runs. It has the possible disadvantage of not reaching a steady-state condition for either type of fuel. This might result in favor of electric heat over gas or oil, since the latter two tend to heat the whole house volume from below while electric heat accounts only for the room in which it is placed. However, such differences were not evident when comparisons were made with later tests.

The tests for House no. 1 from October to December of 1956 were run one week on oil and one week on electricity. The same schedule was again used for the tests in 1957 and 1958, except that from October to the middle of November meter readings for both oil and electric heating were taken daily. For the remainder of the time the readings were again taken weekly.

The average daily or weekly temperatures, as well as the degree-day values, were calculated by using the local Weather Bureau's hourly readings over that period. The degree days were based on the difference between the average temperature for that period and 65 degrees, times the number of days in the period.

To obtain evidence of the extent to which the two types of heating differed in distributing heat in the various parts of the house, a 12-point thermocouple recorder was installed in House no. 1. Thermocouples were placed near the floor, at living height, and near the ceiling in the living room, the kitchen, and one bedroom. The thermocouples were used in the basement, one near the ceiling and the other near the floor. The remaining thermocouple was placed in the outside air about 3 feet from the north wall and well above the ground. Temperatures were recorded over 2 months during the autumn of 1956.

Results from Tests

Detailed figures on tests of this type become quite voluminous and will not be fully given here. Table I gives the essential data collected and processed for five typical weeks on House no. 1. As should be expected, the results vary considerably, and averages taken over the whole test are of greatest value. Such results will be given later in Table II.

Fig. 1 shows a plot of some of the temperatures taken by means of the temperature recorder. The portions selected are typical of one oil and two electric heating periods. The temper-

Table II. Heating Factors from Two Houses Under Test as Well as from Other Sources

Source of Information	Heat Units per Degree-Day: Per 1,000 Square Feet: Oil	Heat Units per Degree-Day: Per 1,000 Square Feet: Electricity	Heat Units per Degree-Day: Per 1,000 Cubic Foot: Oil	Heat Units per Degree-Day: Per 1,000 Cubic Foot: Electricity	Kwhr per Degree-Day: Per 1,000 Square Feet: Oil	Kwhr per Degree-Day: Per 1,000 Square Feet: Electricity	Kwhr per Degree-Day: Per 1,000 Cubic Feet: Oil	Kwhr per Degree-Day: Per 1,000 Cubic Feet: Electricity
House no. 1								
Lowest	1.91	0.586	0.240	0.074	5.58	1.72	0.702	0.216
Highest	2.66	0.917	0.335	0.155	7.78	2.69	0.980	0.338
Average (all test values)	2.23	0.813	0.280	0.103	6.52	2.38	0.820	0.300
House no. 2	Gas		Gas		Gas		Gas	
Lowest	2.76	1.07	0.353	0.137	8.08	3.15	1.03	0.401
Highest	3.92	1.54	0.502	0.197	11.47	4.51	1.47	0.577
Average (all test values)	3.27	1.28	0.417	0.164	9.56	3.75	1.22	0.478
Sharp and Opdenweyer[1]								
Using unit electric heaters (average from 28 houses)								0.300
Using central electric furnaces (average from 21 houses)								0.375
Bush and Woodward[2]								
Averages from 29 houses in Detroit, Mich., area:								
Unit heaters (coldest weekday)								0.24
Unit heaters (average weekday, coldest week)								0.27
Unit heaters (for annual period)								0.26
House in Portland, Oreg.								0.276
From Longview, Wash., tests (average from 38 residences)								0.252
From houses heated with heat pumps: three in Seattle, two in Spokane, two in Portland,								
Highest value obtained								0.310
Lowest value obtained								0.037
Average of all tests								0.150
From 22 houses heated with heat pumps all in or near Boise, Idaho								
For medium-size houses								0.125
For large-size houses								0.146

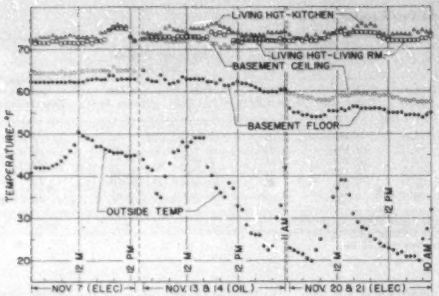


Fig. 1. Temperatures recorded during periods of oil and electric heating

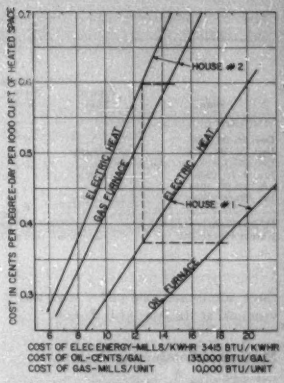


Fig. 2. Cost per heating factor unit when using electric as compared to central furnace for the two houses tested

atures at living height in the living room and kitchen are very nearly the same for both types of heating. None are plotted for the bedroom since these were very similar to those in the rooms shown, except at night when open windows often produced much lower temperatures at times, especially near the ceiling. In general, the floor temperatures were 2 to 4 degrees lower, and ceiling temperatures were 1 to 2 degrees higher than those at living height. When the house was heated with oil the floor temperature tended to be a little higher than with electric heat, probably because some heat was conducted through from below. Likewise, the ceiling temperatures tended to be a little higher with oil, probably because the oil heat was convection in type while the electric was more of the radiant type. However, the differences indicated at both floor and ceiling were small and in general the comfort of living conditions in these rooms was almost identical for the two types of heating.

The greatest difference was noticed in the basement ceiling temperature as should be expected. The basement floor temperature remained fairly stable and continued to drop slowly as the outside temperature decreased regardless of type of heat used.

To abbreviate and present the over-all results of these tests so that they would be of maximum value to the reader regardless of the size of house considered, the local weather conditions, or costs of various types of fuel to be compared, the results are given in terms of heat units per degree-day per 1,000 cubic feet of heated volume where one heat unit is equal to 10,000 Btu (British thermal units). A second set is given in heat units per degree-day per 1,000 square feet of house heated. Other factors which are more commonly used in results from electric heating are kwhr per degree-day per 1,000 cubic feet, as well as for 1,000 square feet. These are also given for oil and gas since there is a definite relation between Btu and kwhr.

Table II lists the average values obtained from all test data available on each of these two houses. The highest and lowest values picked from the test results probably have little significance except to indicate limiting values observed. In both, the conditions are probably abnormal. The low usually occurred early in the autumn, or at times when the family was not making normal use of the house for some considerable period of time. Doubtless the highs were similarly abnormal. The number of these unusual values was comparatively small. The table also gives several similar average values taken from other available test results.

The kwhr per degree-day per 1,000 cubic feet, often termed the heating factor, cannot be considered an accurate measure of heating required for any given house but it appears to be the best general figure available thus far. The heating factor will be low if the house is well insulated thermally, and is well constructed. The factor will likewise be altered by the family living habits. A somewhat lower factor will result for a family which conserves heat by avoiding unnecessary opening of doors and windows in cool weather, or for the family which does not require unusually warm rooms to be comfortable. From the various values given in Table II the kwhr per degree-day per 1,000 cubic feet varies from about 0.25 to 0.45 with the average somewhat below 0.30.

Perhaps some actual cost comparisons would be of more interest to the average reader. For this purpose Table III was computed, using the three schedules of heating energy costs per unit of energy which are shown in the table. Each column lists the cost in cents per degree day per 1,000 cubic feet using the unit-energy cost schedule indicated at the head of the column, and using factors found for each of these two houses from Table II. To compare the results of using one type of fuel the cost of which is given in one schedule with another type given in another schedule, or for costs intermediate between schedules, inter-

Table III. Calculated Fuel Cost Factors*

Location	Fuel Used	Cost in Cents per Degree-Day: Per 1,000 Square Feet at Schedule* A	B	C	Per 1,000 Cubic Feet at Schedule* A	B	C
House no. 1	Oil	2.48	2.89	3.30	0.312	0.364	0.415
	Electricity	2.13	2.80	3.55	0.268	0.352	0.448
House no. 2	Gas	2.98	3.97	5.28	0.333	0.500	0.665
	Electricity	3.40	4.55	5.66	0.428	0.572	0.713

* Using average heat unit factors from Table II and the following schedules of fuel costs:
Schedule A—Oil, 15 cents per gallon; electricity, 0.9 cent per kwhr; gas, 0.8 cent per unit
Schedule B—Oil, 17.5 cents per gallon; electricity, 1.2 cents per kwhr; gas, 1.2 cents per unit
Schedule C—Oil, 20 cents per gallon; electricity, 1.5 cents per kwhr; gas, 1.6 cents per unit

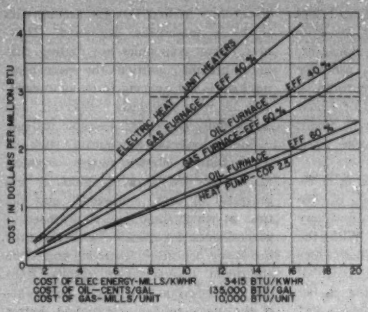


Fig. 3. Calculated cost per million Btu for heating with electric energy as compared to using a central furnace operating at various efficiencies

polations between columns must be made.

To obviate this the results are shown in curve form in Fig. 2, where the cost in cents per degree-day per 1,000 cubic feet plotted against the unit energy costs for the various types of fuel. Thus, to find what price could be paid for electric energy to be the equivalent of oil at 18 cents per gallon in House no. 1, follow the dashed line from the intersection of the oil furnace curve with oil cost at 18 cents per gallon horizontally across to the electric heat curve. The cost is found to be 1.26 cents per kwhr. Since the factor for House no. 2 was higher than for House no. 1, its heating cost using electric energy at this same rate would be 64% higher than for House no. 1. Again following horizontally across to the gas furnace curve, one sees that the cost of gas could be 1.44 cents per unit to be equivalent with electric energy at 1.26 cents per kwhr in House no. 2.

Fig. 3 presents another set of curves for comparing fuel costs. These curves were calculated by using varying costs of fuel energy, indicated on abscissas, to supply a million Btu of heat, with the overall cost in dollars plotted as ordinates. Since all of the heat generated by electric energy is confined to the room containing the electric heaters, this type of heat might be considered 100% efficient. There is, therefore, only a single curve. Furnaces using oil or gas will deliver to the heated space only a portion of the total heat contained in the fuel burned. Only curves for 40 to 60% efficiency are given because the effectiveness of their operation in these tests appears to occur within this range.

Illustrations of the use of these curves may be useful. Electric energy at one cent per kwhr is equivalent to gas heat at 1.17 cents per unit at a furnace efficiency of 40%, or 1.74 cents per unit at 60%. Again, the one cent electric energy is equivalent to oil at 15.7 cents per gallon at a furnace efficiency of 40%, or 23.4 cents per gallon at 60%. From curves in Fig. 2 and Fig. 3 many comparisons can be made, and with the proper factors assumed or determined, a reasonably accurate estimate of the yearly fuel cost for any given house in any given climate can be made.

Apparent Furnace Efficiency

While the efficiency of a furnace can be defined as the ratio of heat delivered to the exhaust plenum to the heat contained in the fuel burned, the ratio of the heat needed in a given house for comfort to its occupants to the heat in the fuel needed for this purpose is almost universally lower than the previously defined efficiency. Even though all the heat produced by electric heaters is available within the enclosure in which they are placed and could be considered 100% efficient, still the amount of heat needed to produce comfort in any part of that enclosure might be much greater with one type of installation than with another.

A very hot surface might heat a small amount of air to a high temperature. The heated air will move to the ceiling where it cools a considerable amount before being displaced to some other part of the room. This will result in poor comfort in either the warm or cool portions of the room. If the same amount of heat is dissipated into a large air volume, the resulting air temperature will be much lower, and it will be displaced to other parts of the room before losing much heat. It will, therefore, give more even temperatures throughout the space.

An alternative to this convection type of heating is the radiant type. Most radiant schemes are partly radiant and partly convective, and in general the amount of heat energy needed in well-designed systems of either type is pretty much the same.

If in these tests the heat needed when heating electrically is considered necessary for comfort, then the apparent efficiency of the central furnaces can be determined.

At the outset of these tests the assumption was made that an oil or gas furnace would be less efficient in early autumn or late in the spring when it is used only a small part of the time, and would be most efficient at the time most heat is needed. Thus an attempt was made to prove or disprove this assumption. The results of this study are given in Table IV.

Two groups of data were selected from the total results obtained: the

Table IV. Efficiency Studies: House No. 1

Degree-Days, Approximately	Heat Units per Degree-Day per 1,000 Cubic Feet Average		Apparent Furnace Efficiency, %
	Oil	Elec.	
From Daily Records			
14/day	0.141	0.092	65
17/day	0.332	0.088	27
20/day	0.337	0.105	31
25/day	0.308	0.131	43
30/day	0.241	0.113	47
33/day	0.276	0.139	60
From Weekly Records			
100/week	0.250	0.078	31
130/week	0.334	0.086	26
150/week	0.335	0.099	30
175/week	0.269	0.105	40
200/week	0.318	0.107	34
270/week	0.300	0.107	36
320/week	0.378	0.109	29
	Average		38

Average Outside Air Temperature, Approximately	Heat Units per Degree-Day per 1,000 Cubic Feet Average		Apparent Furnace Efficiency, %
	Oil	Elec.	
50	0.253	0.092	36
47	0.306	0.096	31
45	0.337	0.086	26
43	0.329	0.109	33
40	0.288	0.151	52
38	0.281	0.130	46
36	0.259	0.099	38
34	0.270	0.116	43
32	0.279	0.139	49
30	0.282	0.123	44
		Average	40

Average from all records taken on House no. 2—39%.

first compared furnace operation with electric for both progressively higher degree-days per day from daily recorded results, and progressively higher degree-days per week from the complete test; the second made a similar comparison with progressively lower average outside air temperatures as the basis of greater heat requirements. The apparent efficiency values found in Table IV seem to indicate there is no noticeable difference in furnace efficiency as the heating demands increase. This seems to check recent reported tests on this comparison made elsewhere.

The apparent furnace efficiency varies considerably in both groups, but the average for each group is near 40%. A similar average value obtained from the month's run on House no. 2 gave 39% as the apparent efficiency even though the heat insulation and living characteristics of House no. 1 and House no. 2 were quite different.

The test efficiency of central furnaces similar to those in either of these two houses is usually assumed to be about 60 to 80%. The wide divergence between the apparent efficiency derived from this type of test and the factory type of test is probably the reason why the usual heating contractor uses a relatively high heat loss factor for a house when proposing a suitable furnace. This difference is well illustrated by figures given in paper by Jones and Linden.[3] Three separate heat loss calculations were obtained on the same house as follows:

A near-by electrical utility	42,000 Btu per hour
Heating cable manufacturer	48,000 Btu per hour
Reliable heating contractor	75,000 Btu per hour

The utility based their results on experience with electrically heated houses on their system. The heating contractor actually calculated 57,000 Btu per hour but recommended 75,000 Btu per hour.

Comments and General Deductions from Tests

The results from tests of this nature depend upon so many variables that only general conclusions can be safely stated. The causes for the decided difference in heating factors computed for these two houses might be as follows: House no. 1 was better heat insulated in both ceilings and windows than was House no. 2. House no. 1 was on a level terrain surrounded to some extent by other houses, while House no. 2 was on a hillside facing Puget Sound and subject to more south and west, and even north winds. The family in House no. 1 was more adult in nature, probably away to school or work for portions of the day, while the occupants of House no. 2 consisted of a child in ill health and two adults probably in the house almost continuously.

Regardless of the differences in house construction or living habits of the occupants, the heat needed from any source is about the same. With the ever-increasing cost of the usual types of fuel, and the stability in cost of electric energy, electric heating is becoming very competitive, and more and more popular. In most cases the lowest cost step in electric rates can be considered the cost price for electric heating since the home owner must pay for the higher cost levels used for light, range, and water heating even though he has no space heating. At present prices of fuel, electric energy at one cent per kwhr, or perhaps even somewhat higher, is competitive with oil or gas in most localities, and when the convenience and cleanliness of electric heat is considered higher kwhr costs are still attractive.

The occupants of House no. 1 were paying 18.7 cents per gallon for oil plus 0.8 cent per gallon for furnace motors, while their lowest step electric rate was one cent per kwhr. Their over-all saving in using electric heat was about 20%.

The argument might be presented that the central furnaces in these two houses were both operating at an abnormally low efficiency. The fact that both showed similar results, and that the furnace in House no. 1 was relatively new and well designed, and apparently operating very normally, strongly supports the assumption that both were quite typical of the normal central-furnace operation.

The author has for some years been collecting cost information on heating houses with oil, as well as with electric heat, and while such comparisons are difficult and no figures will be given, the general results appear to compare quite favorably with the results of these tests.

Residential Heating as a Load for the Power Companies

For many years the power companies considered that electric space heating resulted in a load factor on the system too low to be desired, except at a rate greater than the minimum in their rate structure. At first thought this appears to be a reasonable deduction. However, many companies are finding that under present conditions the space heating load is not proving to be as undesirable as was previously suspected, and occasionally is proving desirable instead.

Since domestic appliances, such as water heaters, electric ranges, washers, and dryers are becoming more or less universal in every household, and the time at which these devices are used in general does not coincide with the time the space heating is greatest, and since additional diversification results from the diversified habits of the occupants of the many residences supplied, better over-all load factors have been obtained than were previously thought possible. Again, in many parts of the United States the sizeable summer air-conditioning load tends to balance the winter heating load.

Some companies are finding the average additional demand of each electrically heated residence is small compared to the average connected capacity of the heating load. One company states that this average additional demand amounts to approximately 2 kw for an electrically heated house. This would appear to be very reasonable in comparison to the total connected kilowatt load of the average house today. Doubtless this condition will become more favorable with the ever-increasing trend of using more electric energy in the average household, and the space-heating load will become a desired rather than a dreaded additional load as was the case with both the electric range, and later the electric water heating load, some years ago.

References

1. Electric-House-Heating Tests in Oregon (abstract). W. L. Sharp, A. E. Opdenweyer. *Electrical Engineering*, vol. 67, Jan. 1948, p. 78.

2. Residential Electric Space Heating in Detroit for 1952–1953 Heating Season. A. E. Bush, R. P. Woodward. *AIEE Transactions*, vol. 74, pt. II, July 1955, pp. 200–04.

3. Heat Pump and Heating Cables Installed in the Same Residence: First Year of Comparative Operation. C. W. Jones, E. E. Linden. *Ibid.*, vol. 77, pt. II, Mar. 1958, pp. 43–48.

Discussion

E. E. Linden (The Narragansett Electric Company, Providence, R.I.): Mr. Smith's paper, which develops a relationship between residential electric heat and heat from fossil fuels, is based upon carefully collected data and should be considered a valuable reference in this field. In this discussion I

raise two points: one concerns the significance of estimating electric heating requirements and costs by classical fuel-fired procedures and the other points up the importance of improving present-day heat pumps.

In estimating heat losses, usually some form or derivative of American Society of Heating and Air Conditioning Engineers Guide procedures and recommendations is used. Assumed efficiency of 70% for oil and 80% for gas, when applied to the calculated loss, result in estimated annual fuel cost, which in most cases turn out to be very close to actual operating figures. Now if this procedure is extended, an estimate for direct conversion electric heating can be made by pricing the kwhrs equivalent to the Btu content of the oil or gas use estimates, times their respective efficiencies. Estimates so made, however, turn out to be far above what the actual cost of operation turns out to be later.

One example might be useful to relate the impropriety of converting fuel-fired estimates to electricity. A residence outside of Providence, R. I., has an estimated fuel-fired heat loss of 56,000 Btu per hour which at 14 cents per gallon for oil results in an annual operating cost estimate of about $200. Converting the Btu content to kwhrs results in an estimate for electricity of almost $800. This residence has been electrically heated and carefully metered for 2 years and found to have a heat loss nearer 33,000 Btu per hour with an operating cost for direct conversion electricity, at 2 cents per kwhr, somewhat under $500. Now, if the efficiency for the oil-fired estimates had been taken as 40%, as found by Mr. Smith, and heat loss at 33,000 Btu per hour, then the operating cost for oil would still be about $200. This, of course, would put electricity and the fuels into a truer relationship.

When a home builder considers comparative heating plants, he does so on a basis of estimated costs. While present procedures seem adequate for fuel-fired systems, it seems apparent that they cannot be extended through simple conversion to price electricity. In the first place the high heat loss estimate will raise the first cost of heating equipment and service and develop an excessive kilowatt demand; and in the second place the kwhr use becomes forbidding. Estimates of first cost and operating costs for an electrically heated residence will be most competitive when based on experience gained from residences presently heated by electricity.

While Mr. Smith does not deal with the heat-pump problem, there is probably as much question involved with machine ratings and coefficient of performance (COP) as there is with the efficiency figures for fuel fired systems. The heat pump has enjoyed much publicity, has broad appeal, but may need a lot more work in development and design before it finally arrives. The much referred to COP of present-day machines is usually represented as 3 to 1, and sometimes 4 to 1. It should be based upon heat output compared to the heat equivalent of the electrical input. While under test at the afore-mentioned Rhode Island residence, a water-source heat pump with a rated output of 42,000 Btu was shown to have an output of under 35,000 Btu per hour. Here is a case where, based on real output, the COP is 2.1; on specified output, 2.6; on calculated heat loss, 3.5. If this holds true of heat pumps in general, then the challenge to the industry is apparent. A suggested goal might be a 40,000-Btu (real output) machine complete with all accessories to sell for $750 installed, a service contract covering the complete equipment for $25 per year, and a heating bill on 2 cent electricity of under $150 per year.

G. S. Smith: I would like to thank Mr. Linden for calling attention again to the existing confusion in the use of the term "furnace efficiency" and for clarifying the difficulty by means of a concrete example from actual experience. By using such reasoning the potential home owner can get a fairly accurate comparison of the cost for various types of heating.

I must also agree with Mr. Linden that the effective heat to produce comfortable living conditions in a home may be less than that calculated by using the conventional COP of a heat-pump installation. Here again a considerable amount of information from actual experience is needed.

Again I agree with a comment made by one of the reviewers of this paper, that this one more or less isolated test does not present sufficient material upon which to base final conclusions. However, the opportunity for careful tests of this type are seldom found, and could be very expensive if carried out by other means than were possible in this case. I do believe that the results given in this paper are of great value, especially since the heating factors derived compare so favorably with those found in several other studies of electric heating. More tests of this type would be most welcome.

Effect of Driving Source Impedance on Motor Performance

ALBERT DIAMOND
ASSOCIATE MEMBER AIEE

TO DATE, little attention has been focused on the resulting over-all system performance due to the effect of driving-source impedance on servomotor characteristics. The torque–speed characteristics of a motor, as furnished by the manufacturer, may be correctly used for calculating the motor transfer function only when the identical test conditions relating to the driving source impedance to both phases are satisfied.

The manufacturer usually obtains his data while operating both phases of the motor from essentially a zero source impedance. Unfortunately, in the usual system application, the control phase of the motor is driven from an amplifier with a finite output impedance, and many times a capacitor is used in series with the fixed winding for phase-shifting purposes. These source impedances will modify the torque–speed characteristics of the motor, often ruining its inherent damping, and thereby cause degeneration of the system performance.

Assumptions

1. For the purpose of simplification, it is assumed that the motor to be discussed has its fixed phase driven from a zero source impedance and the control winding from an amplifier with a capacitor across the control winding for power factor correction.
2. The leakage reactance of the rotor is negligible. This is particularly true of a servomotor since the rotor resistance is high.
3. The core loss is negligible.

Discussion

It is the object of this section to show how resonating the amplifier output and motor control winding in order to obtain higher amplifier efficiency affects motor performance in a system.

Chang[1] has shown that the servomotor with zero voltage on the control field does not develop a positive torque at any speed provided that R_s, the effective rotor resistance, is greater than $|Z|$ where $|Z|$ is defined as the Thévenin equivalent impedance the rotor sees looking toward the amplifier. Conse-

Paper 58-1028, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted June 25, 1957; made available for printing July 30, 1958.

ALBERT DIAMOND, formerly with Helipot Corporation, Newport Beach, Calif., is now with United Aircraft Corporation, Comack, N. Y.

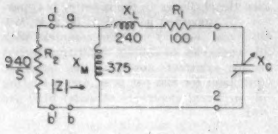


Fig. 1. Per-phase equivalent circuit of a servomotor with a capacitor across the winding

quently, neglecting friction and windage, single phasing will occur when $R_2 < |Z|$. Also, for values of R_2 less than, but of the order of, $|Z|$, poor damping qualities particularly near zero speed will be exhibited.

Equivalent Circuit

Fig. 1 is the equivalent circuit representation of the control phase of the servomotor. As a limiting case, it is assumed that the output impedance of the amplifier is purely capacitive. This capacitor is used to resonate the amplifier and motor and is represented as X_c across terminals *1* and *2* of the motor. It is further assumed that the amplifier is not delivering any voltage to the motor.

From Fig. 1, the impedance $|Z|$ looking into terminals *a-b* is given as

$$|Z| = \left\{ \frac{X_m{}^2 R_1{}^2 + X_m{}^2 [R_1{}^2 + (X_L - X_c) \times (X_L - X_c + X_m)]^2}{R_1{}^2 + (X_L + X_m - X_c)^2} \right\}^{\frac{1}{2}} \quad (1)$$

To compare analytical and experimental results, the per-phase equivalcnt circuit of a Beckman servomotor type *15SM460* was obtained. The value of the parameters were for 400-cycle-per-second operation.

R_1 = resistance of control winding = 100 ohms
X_L = leakage reactance of control winding = 240 ohms
X_m = mutual reactance = 375 ohms
R_2 = effective rotor resistance = 940 ohms

Substitution of the foregoing values in equation 1 resulted in the data found in Appendix I. These data were plotted and the ensuing graph is as shown in Fig. 2. From Fig. 2, it is seen that $|Z| > R_2$ from $525 \leftarrow X_c \rightarrow 840$. Thus the motor should single phase for resonating capacitors between 0.76 and 0.48 microfarad. This fact was verified experimentally on the Beckman motor. It is particularly significant that this motor would run single phase because its inherent design parameter of $R_2/X_m = 2.5$ would result in very adequate damping even when driven from an infinite source impedance of noncapacitive elements.

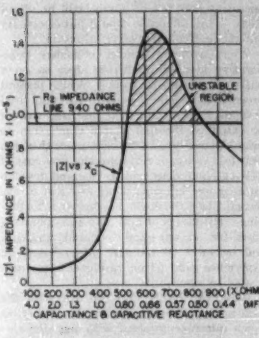


Fig. 2. Comparison of effective rotor resistance with the impedance that it sees versus capacitance across the winding

Rapid Evaluation

It is important to know whether a particular motor can be made to run single phase for a particular value of capacitance across the winding.

To determine the maximum value of $|Z|$ versus X_c, equation 1 is differentiated. The value for X_c which causes $|Z|$ maximum is given as:

$$X_c = X_L + X_m + \frac{-X_m + \sqrt{X_m{}^2 + 4R_1{}^2}}{2} \quad (2)$$

while

$$|Z|_{max} = \frac{2X_m R_1}{(X_m{}^2 + 4R_1{}^2)^{\frac{1}{2}} - X_m} \quad (3)$$

Thus, by substitution of the motor parameters in equation 3, a quick check of whether or not a particular servomotor can be made to single phase is readily apparent.

Modification of $|Z|_{max}$

In the event the driving amplifier has a resistive output impedance R_3 which is not negligible with respect to X_c the expression for $|Z|_{max}$ must be further modified. Fig. 3 represents the equivalent circuit of the motor where X_c is now shunted by R_3. R_3 represents the

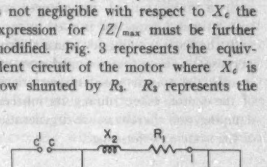


Fig. 3. Per-phase equivalent circuit of a servomotor with a resistor and capacitor across the input simulating the driving source impedance

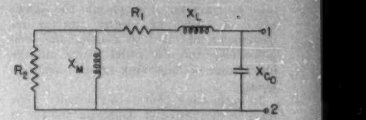


Fig. 4. Per-phase equivalent circuit of a servomotor with a capacitor across the input terminals resonating it for unity power factor

output impedance of the amplifier. The modified $|Z|$, now called $|Z'|$, looking into terminals *c* and *d* is now

$$|Z'| = \left\{ \frac{X_m{}^2 R^2 + X_m{}^2 [R^2 + (X_L - X) \times (X_L + X_m - X)]^2}{R^2 + (X_L + X_m - X)^2} \right\}^{\frac{1}{2}} \quad (4)$$

where

$$R = \frac{R_1 + R_3 X_c{}^2}{R_3{}^2 + X_c{}^2} \quad (5)$$

$$X = \frac{R_3{}^2 X_c}{R_3{}^2 + X_c{}^2} \quad (6)$$

The relationship between the values of capacitors which will cause a single-phasing condition and the capacitors needed for unity power factor also has to be considered. This is a prime factor because it is conceivable that, for any given motor, the value of capacitor needed for unity power factor will not fall inside of the range of capacitor values which will result in single phasing.

For a first-order approximation, it was assumed that the capacitor needed to resonate the motor winding alone would be the same as the value needed to resonate the winding driven from an amplifier of a high source impedance.

Fig. 4 represents the equivalent circuit and X_{co} is given as

$$X_{co} = \frac{X_L + R_1{}^2 (R_2{}^2 + X_m{}^2) + X_m R_2 (2R_1 X_m + X_L R_2)}{X_L (R_2{}^2 + X_m{}^2) + X_m R_2} + \frac{X_m{}^2 R_2{}^2 (X_m + R_2)}{[X_L (R_2{}^2 + X_m{}^2) + X_m R_2][R_2{}^2 + X_m{}^2]} \quad (7)$$

Substitution of the parameter values (taken from the Beckman motor) in equation 7 results in $X_{co} = 615$ ohms ($C = 0.65$ microfarad). This indicates that if unity power factor correction were attempted using no corrective measures, single phasing would occur.

To verify the foregoing results experimentally, a breadboard amplifier was built and the motor resonated for unity power factor. The measured value of capacitor needed for unity power factor was 0.68 microfarad (590 ohms) and the motor was verified to run single phase.

This condition was corrected by using negative feedback around the amplifier loop. Enough negative feedback was inserted (keeping amplifier gain constant) until the corner frequency of the motor and amplifier, in combination, yielded essentially the same value as the motor itself driven from a zero source impedance.

Corrective Measures

When the desire to correct for power factor causes $|Z|$ or $|Z'|$ to be greater than R_2, resulting in poor system response or even single phasing, it is the responsibility of the amplifier designer to take corrective measures. Some of the obvious methods would be

1. Add a resistance in series with control winding of motor (same effect as increasing R_1).

2. Shunt resistance across resonating capacitor.

3. Reduce output impedance of amplifier by employing negative voltage feedback around amplifier loop.

Methods 1 and 2 are both wasteful of amplifier power output. Method 3 sacrifices voltage gain, usually a more reasonable compromise, and is recommended by the author.

Conclusions

The results indicate that greater caution should be exercised when matching a motor and amplifier. Considerable sacrifices in dynamic system performance can result if the full capabilities of the servomotor are not realized due to improper driving arrangements or circuitry.

Appendix I. Derivation of $|Z|_{max}$

$$|Z| = \frac{\{X_m^4R_1^2+X_m^2[R_1^2+(X_L-X_c)\times(X_L-X_c+X_m)^2]\}^{\frac{1}{2}}}{R_1^2+(X_L+X_m-X_c)^2} \tag{8}$$

$$\frac{d|Z|}{dX_c} = [R_1^2+(X_L+X_m-X_c)^2]\times[R_1^2+(X_L-X_c)(X_L-X_c+X_m)]\times[-2X_L+2X_c-X_m]-[X_m^2R_1^2+[R_1^2+(X_L-X_c)(X_L-X_c+X_m)]^2]\times[-2X_L+2X_c-2X_m] \tag{9}$$

Let

$$X_c = X_1+X_m+Q$$

and set

$$\frac{d|Z|}{dX_c} = 0$$

$$Q^4+X_mQ^3+X_mR_1^2Q-R_1^4=0$$

Let

$$a = X_m$$

and

$$b = R_1^2$$

and by completing the squares

$$Q^2+\frac{a}{2}Q = \pm\left(\frac{a}{2}Q-b\right)$$

then

$$Q_1 = jR_1$$

$$Q_2 = -jR_1$$

$$Q_3 = \frac{-XM+\sqrt{X_m^2+4R_1^2}}{2}$$

$$Q_4 = \frac{-X_m-\sqrt{X_m^2+4R_1^2}}{2}$$

Q_1 and Q_2 are trivial roots, therefore,

$$X_c = X_L+X_m-\frac{X_m+\sqrt{X_m^2+4R_1^2}}{2} \tag{10}$$

gives $|Z|_{max}$ and

$$X_c = X_L+X_m+\frac{-X_m-\sqrt{X_m^2+4R_1^2}}{2} \tag{11}$$

give $|Z|_{min}$. The actual value of $|Z|_{max}$ is obtained by substituting equation 10 into equation 8.

Appendix II. Derivation of X_{co} Capacitive Reactance Needed for Unity Power Factor

Looking into terminals *1* and *2* of Fig. 4, the total impedance is given as:

$$Z = \frac{\dfrac{X_c^2X_m^2R_2}{R_2^2+X_m^2}+X_c^2R_1}{D} - j\left(\frac{\dfrac{X_cX_m^2R_2^2+X_cX_m^2R_1^2}{(R_2^2+X_m^2)^2}+\dfrac{2X_cX_m^2R_2R_1+2X_cX_LX_mR_2^2-\dfrac{X_c^2X_mR_2^2}{R_2^2+X_m^2}}{D}}{D}\right) - j\left(\frac{X_cR_1^2+X_cX_L^2-X_c^2X_L}{D}\right)$$

where

$$D = \left(\frac{X_m^2R_2}{R_2^2+X_m^2}+R_1\right)^2+\left(\frac{X_mR_2^2}{R^2+X_m^2}+X_L-X_c\right)^2$$

For unity power factor, imaginary term must equal zero. By multiplying the numerator by $(R^2+X_m^2)^2$, and extracting the root $X_c = 0$, after algebraic manipulation the following is obtained:

$$X_c = \frac{X_L+R_1^2(R_2^2+X_m^2)+\dfrac{X_mR_2(2R_2^2X_m+X_LR_2)}{X_L(R_2^2+X_m^2)+X_mR_2^2}+\dfrac{X_m^2R_2^2(X_m+R_2)}{[X_L(R_2^2+X_m^2)+X_mR_2^2][R_2^2+X_m^2]}}{}$$

References

1. The Equivalent Circuit of the Capacitor Motor, S. S. L. Chang. *AIEE Transactions*, vol. 66, 1947, pp. 631–40.

2. Operating Characteristics of 2-Phase Servo Motors, R. J. W. Koopman. *Ibid.*, vol. 68, 1949, pp. 319–29.

3. Transfer Function for a 2-Phase Induction Servo Motor, Lloyd O. Brown, Jr. *Ibid.*, vol. 70, 1951, pp. 1890–99.

Discussion

R. M. Saunders (University of California, Berkeley, Calif.): The author brings out points in the application of a-c servomotors which are seldom mentioned in manufacturers' literature. In fact, it is often difficult to elicit either from supplied data or scientific deduction just how or under what conditions manufacturers obtained their data.

As the author indicates, zero source impedance and exactly 2-phase voltage conditions are generally those under which a motor is tested by the manufacturer. In application a common mode is to have a 2-phase source with the reference phase supplied by a zero impedance source and the control phase with a substantial impedance (either actual or Thévenin equivalent) in series with its source. Another very common application not touched upon by the author is to excite both phases from a single-phase source, the control phase being excited as described by the author and the reference phase having a series capacitor to obtain approximately 90-degree phase shift between the reference phase and the control phase currents. It would be well if the author extended his development in a subsequent paper to include the single phase case.

Fig. 1 apparently is not correct as it stands in that the whole negative sequence circuit and the contribution of the reference phase voltage to the positive sequence circuit are not shown. On the other hand, it may be that the author is using a different terminology from that generally accepted by motor theorists. Hence it would be of interest for the author to explain how he gets his Fig. 1 from the generally accepted servomotor equivalent circuit given by Fig. 7 of Koopman (see reference 2 of the paper).

It should be pointed out that voltage gain is of no significance in the power-modulation stage of a control system. What really counts is power gain or the measure of system's ability to control or modulate the flow of large blocks of power or energy with relatively small input power.

Motor-Operated Switch for Space Heating Control

J. C. BECKETT
MEMBER AIEE

Synopsis: Resistance load (electric heaters, incandescent lights, etc.) can be switched most satisfactorily by a relatively slow moving motor-operated type of switch. Noise, contact wear, and general maintenance are all less than with the ordinary type of magnetic contactor, which is designed for switching inductive loads.

POOR OPERATION of ordinary magnetic contactors led, several years ago, to a new design of motor-operated switch for electric-heating circuits. Although the new switch was originally intended to eliminate special difficulties that arose when switching electric-heater load from one voltage to another, the motor-operated switch proved to give superior performance for all kinds of resistance load switching.

Fig. 1 shows a 3-pole single-throw motor-operated switch used in 240-volt and 480-volt a-c circuits for on-off switching of resistance loads. In Fig. 2 is shown a battery of motor-operated switches in an industrial installation: 12 100-ampere 240-volt 3-phase motor-operated switches control a 450-kw industrial prepaint washer and bake oven.

The double-pole double-throw switch designed to switch electric-heating load from full voltage to half voltage is shown in Fig. 3. This switch, used in load regulators, can be actuated either by a current relay, an outdoor thermostat, or manually.

Electric Heating Load Regulator

The control of space heating is complicated by the changing seasons and weather conditions. A device for changing voltage in order to control heating-system capacity is now recognized as an essential component of a good heating system. In cold weather, when the full capacity of the heating system is needed, the heaters (each thermostatically controlled) are connected by the load regulator to the incoming 240-volt line. In mild weather, when the outdoor temperature is higher, the full capacity of the system is not required and closer temperature regulation can be maintained in the building by connecting the heaters to a 120-volt supply.

In addition, rate structures often penalize high peak loads. By reducing the voltage on the electric-heating system when there is high demand for other purposes (such as electric ranges, clothes dryers, water heaters, and various appliances), the peak demand, and hence the utility charges, can be kept to a minimum. This does not usually result in any failure to maintain proper temperature in the house, for some or all of the heat produced by the range or other appliance is likely to be retained within the house, and when these are in operation the full capacity of the heaters is not needed.

The necessary device must disconnect load from one voltage and connect it to another rather quickly. Since household heating systems range in capacity from about 6 to 30 kw, usually at 240 volts a-c, the device was designed to operate in this range. A schematic wiring diagram is shown in Fig. 4. The electric-heater load can be connected either to 240 volts (line to line) or to 120 volts (line to neutral).

Solenoid Operation

Initially, a standard magnetically operated double-throw switch was used, but it was observed that there was occasionally excessive arcing at the contacts, resulting in considerable deterioration of the silver surfaces. Various schemes for controlling this excessive arcing were tried but none proved completely successful.

Oscillographic tests gave a clear picture of what was happening. When the load regulator operated, the arc drawn between the opening contacts was sometimes not extinguished before the closing contacts came together. The result was a virtual short circuit from line to neutral of the supply, through the arc. The heavy short-circuit current damaged the contact surface before it could be stopped; the duration of the arc varied from a 1/2-cycle to many cycles.

Fig. 5(A) shows a successful operation of a load regulator equipped with a solenoid-operated switch. The two upper traces show voltage (and hence current) to the two resistance loads. The lower trace shows current through the current relay in the appliance circuit that operates the load regulator (see Fig. 4). In this instance, the 240-volt circuit opened as the voltage approached zero, an ideal condition. The arc was extinguished at the time marked *arc clears*. Contacts closed on the 920-volt supply in less than a 1/2 cycle, and normal current (except for minor irregularities resulting from bounce of the contacts) is shown thereafter.

Fig. 5(B) shows another successful operation of the solenoid-operated switch. Contacts started to open at the peak of

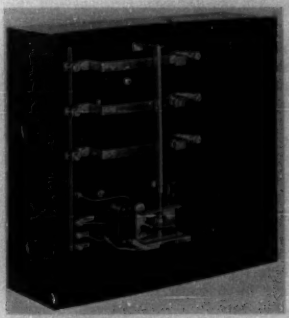


Fig. 1. Motor-operated switch used on 240- and 480-volt a-c circuits for noninductive load

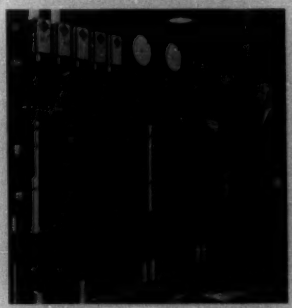


Fig. 2. Twelve 100-ampere 240-volt 3-phase motor-operated switches controlling 450-kw industrial prepaint washer and bake oven

Paper 58-1014, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19–22, 1958. Manuscript submitted May 9, 1958; made available for printing May 20, 1958.

J. C. BECKETT is with the Wesix Electric Heater Company, San Francisco, Calif.

Fig. 3 (left). Motor-operated double-throw switch used to switch voltage on heater circuits

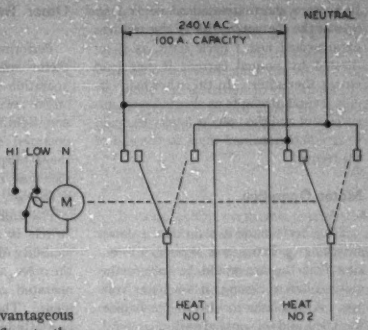


Fig. 4 (right). Wiring diagram of motor-operated double-throw switch shown in Fig. 3

the voltage wave but current was not interrupted until the applied voltage reached zero, at the time marked *arc clears*. Contacts making connection to the 120-volt circuit closed only about 1/6 of a cycle later.

About once in 30 operations there was a short circuit. An example is seen in Fig. 5(C). The contacts started to open just after a zero of the load current and voltage. This is the least advantageous time. Current continued to flow to the heater load through the arc that was being drawn out in the contactor, and at the time marked *contacts close* the 120-volt contacts closed without the arc having been extinguished. Both line wires were simultaneously short-circuited to neutral, and voltage on all three circuits was pulled down practically to zero. Several jagged traces show instants of near extinction of the arcs, but these were unsuccessful, and by the end of the oscillogram the arc was so firmly established that all line voltages were held down to zero.

Fig. 5(D) shows a partial failure. At the instant marked *contacts close*, both

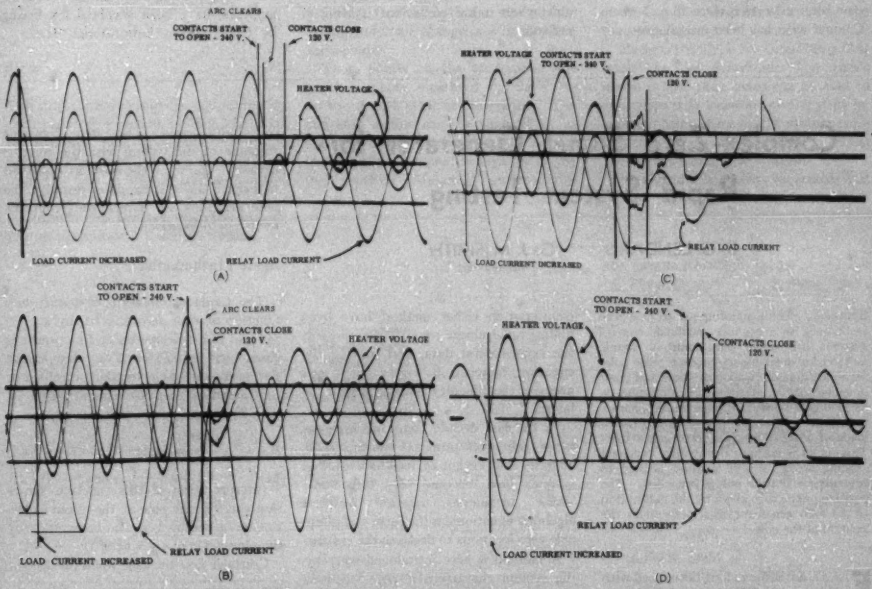


Fig. 5. Oscillograms showing double-throw switching from 240 to 120 volts a-c with conventional magnetic double-throw switch

A—Successful voltage switching of two heater circuits, voltage near zero on opening
B—Same as (A) except voltage near peak on opening
C—Unsuccessful voltage switching of two heater circuits showing arc over on both circuits
D—Same as (C) except arc over continued on one heater circuit only

lines were short-circuited to neutral by arcs at the contacts. One of the arcs was extinguished just 1 cycle later, as is indicated by normal current in one load circuit thereafter. In the other load circuit several attempts at clearing were unsuccessful and the voltage finally dropped to zero and remained there to the end of the record.

Motor Operation

These tests made it plain that a slower switching operation was necessary to ensure that the arc would be fully extinguished before closing on the lower voltage. A simple motor-driven mechanism was tried, and this proved quite successful. As shown in Fig. 3, a small 30-watt shaded-pole motor is used to drive an eccentric arrangement for operating the contactor. (Contact points are at the top of the picture to the right of middle.) The time required to switch from 240 to 120 volts is approximately 1/3 second (about 12 cycles); this gives ample time for an arc to be fully extinguished.

These motor-operated switches have now been in service more than 7 years. Contact wear has been insignificant.

Other Switching Applications

Performance of the motor-operated contactor was so satisfactory that motor operation has been applied to simple on-off switching of industrial heating and lighting loads. Although solenoid-operated contactors could be used for these purposes, the motor-operated device has marked advantages in that it is quieter, and needs less maintenance.

The annoying "bang" of the magnetic switch is absent. Considering the desirability of minimizing noise in the home, the office, and the school, the quiet motor-operated contactor is a great improvement. The motor-operated switch can be located within the space-heater cabinet which is more convenient and economical than placing it in a remote location. Large-capacity single-unit electric space heaters, such as ceiling-suspended units and cabinet heaters, are now in service in schools with motor-operated switches installed as part of the heater. Lighting loads with motor-operated switches are also in successful use; quiet operation and low maintenance make this switch desirable where relay control of lighting is necessary.

Fig. 1 shows such a switch. It is operated by a reinforced-plastic eccentric bar actuated by the motor. This bar exerts pressure against spring contact arms to open the circuit. The spring pressure alone maintains adequate contact pressure. This design minimizes any alignment problems and ensures that there will be positive force to open the circuit. Limit switches to control the actuating motor have wiping contacts, avoiding trouble from dirt and oxidation. Several switches may be operated in parallel from a single control device such as a thermostat or clock switch, or in sequence by means of a separate set of limit switches (Fig. 2). This type of switch requires energy to open as well as to close, requiring a thermostat or other device with double-throw contacts, but introduces no other complications.

Maintenance is practically negligible. The contacts of such a switch have required no attention in more than 200,000 cycles of operation. No arc-extinguishing devices are employed.

Two-pole and 3-pole devices rated at 240 volts, 100 amperes, and 480 volts, 70 amperes, have been accepted for listing by Underwriters' Laboratories, Inc.

Complex-Zero Signal Generator for Rapid System Testing

G. G. LENDARIS
ASSOCIATE MEMBER AIEE

O. J. M. SMITH
MEMBER AIEE

Synopsis: The expanding use of pole-zero techniques in designing feedback control systems has made paramount a simple method for determining characteristic pole-zero locations of existing systems and components. As a step in this direction, a method is presented for determining characteristic pole locations of a system. Use is made of a signal from a "zero generator." This signal is fed into the system to cancel existing poles. A null method is used to determine when the pole is cancelled. The position of the zero given by the calibration of the zero generator then determines the location of the pole.

ENGINEERS have long been faced with the problem of finding an analytical description of a physical system. To date the methods of analysis have been confined to steady-state frequency response analyses and transient response analyses.[1-3] The chief problems encountered in either method have been finding an algebraic equation to represent the experimental data, and, having the equation, factoring it. Everyone is well aware of the difficulty associated with the latter.

In the last decade the powerful root-locus techniques proposed by Evans for analysis and design of feedback control systems have become very important. These techniques require factored algebraic equations in the form of s-plane pole-zero locations to describe the system.

A method is here introduced whereby the system characteristic pole locations can be found directly in the laboratory, thus by-passing the laborious process of synthesizing algebraic equations from experimental curves, and then factoring them.

The method involves using a signal containing zeros. The step transient response of the system with the zeros is obtained, and when the zeros are observed to cancel the system poles exactly, the pole locations are determined from the known zero locations.

Some Mathematics

The Laplace transform of a unit step function at time zero is given by $\mathcal{L}U(0) = 1/s$. For a delayed step function it is given as $\mathcal{L}U(T) = e^{-sT}/s$. An initial step of magnitude a, to which is added or subtracted a step of magnitude b at time T, has a Laplace transform

$$F(s) = \frac{a}{s} \pm \frac{be^{-sT}}{s} = \frac{1}{s}(a \pm be^{-sT}) \qquad (1)$$

Interpretation of this function shows that there is one pole at the origin of the s-plane (given by $1/s$) and many zeros at the places where $a = \mp be^{-sT}$.

Consider first the case when the second

Paper 58-1027, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19-22, 1958. Manuscript submitted January 16, 1958; made available for printing June 20, 1958.

G. G. LENDARIS and O. J. M. SMITH are with the University of California, Berkeley, Calif.

step of magnitude b is subtracted. Then

$$a=+be^{-sT} \tag{2}$$

$$e^{sT}=\frac{b}{a} \tag{3}$$

s is a variable in the s-plane whose coordinate axes are σ and $j\omega$. The roots of equation 3 give specific values to s and will be denoted by $s=\alpha+j\beta$. Then

$$e^{sT}=e^{(\alpha+j\beta)T}$$

$$=e^{\alpha T}(\cos\beta T+j\sin\beta T)=\frac{b}{a} \tag{4}$$

Since b/a is all real, $\sin\beta T$ must be zero and $\cos\beta T$ must have a magnitude of unity. Since b/a is positive, $\cos\beta T$ must be plus one. Therefore,

$$\beta T=\pm 2n\pi \tag{5}$$

where n is any positive integer, $n=0,1,2\ldots$ This fulfills the conditions:

$$\sin(\beta T)=0 \text{ and } \cos(\beta T)=+1 \tag{6}$$

Then

$$|e^{sT}|=e^{\alpha T}=\frac{b}{a} \tag{7}$$

$$\alpha=\frac{1}{T}\ln\frac{b}{a} \tag{8}$$

$$\beta=\pm\frac{2n\pi}{T} \tag{9}$$

This first condition may be represented in the s-plane by a zero at $\alpha=1/T\ln b/a$ on the real axis, plus an infinite column of zeros at the same α but each separated from the next in the $\pm j\omega$ direction by $2\pi/T$.

Consider next the case when the second step of magnitude b is added. Then,

$$a=-be^{-sT} \tag{10}$$

$$e^{sT}=e^{\alpha T}(\cos\beta T+j\sin\beta T)=-\frac{b}{a} \tag{11}$$

Since $-b/a$ is all real, $\sin(\beta T)$ must be zero again, and $\cos\beta T$ must have unity magnitude. Since $-b/a$ is negative, $\cos\beta T$ must be minus one. Therefore,

$$\beta T=\pm(2n+1)\pi \quad n=0,1,2\ldots \tag{12}$$

This fulfills the conditions:

$$\sin(\beta T)=0 \text{ and } \cos(\beta T)=-1 \tag{13}$$

Then

$$-|e^{sT}|=-e^{\alpha T}=-\frac{b}{a} \tag{14}$$

$$\alpha=\frac{1}{T}\ln\frac{b}{a} \tag{15}$$

$$\beta=\pm\frac{(2n+1)\pi}{T} \tag{16}$$

This second condition may be represented by one pair of complex zeros at

$$\alpha=\frac{1}{T}\ln\frac{b}{a},\quad \beta=\mp\frac{\pi}{T}$$

with an infinite column of zeros again each separated from the next by $2\pi/T$.

It becomes cumbersome to speak of an infinite column of zeros, therefore the following notation will be used:

The transformation $z=e^{sT}$ offers the advantage that in the z-plane the infinite number of s-plane zeros are represented by one z-plane zero; this is accomplished by a corkscrew-type folding in the conformal transformation. This can be explained as follows:

By definition

$$z=e^{sT}=e^{\sigma T}e^{j\omega T} \tag{17}$$

It is noted that

$$|z|=e^{\sigma T} \tag{18}$$

and

$$\underline{/z}=\omega T \tag{19}$$

Therefore, a constant σ line in the s-plane transforms into a trajectory in the z-plane whose radius is $e^{\sigma T}$ (constant) and angle is βT (β is the variable) hence a circle. One revolution in the s-plane corresponds to a change in ω by $2\pi/T$ in the s-plane.

Thus, points on the constant σ line in the s-plane separated by $2\pi/T$ lie on top of each other in the z-plane. The constant σ line can be thought of as spiraling upward around an axis normal to the z-plane, passing through the origin, at a constant distance $e^{\sigma T}$ and variable angle ωT. Hence the vertical projection of all points separated by $2\pi/T$ will coincide to give one point in the z-plane. One of these spirals can be drawn for every σ, hence the extension to the model of a plane spiraling upward about the z-plane origin as an axis. Note for

$$\begin{aligned} &\sigma=+\infty && |z|=\infty \\ &\sigma=0 && |z|=1 \\ &\sigma=-\infty && |z|=0 \end{aligned} \tag{20}$$

Because of this corkscrew-type distortion, it is not convenient to correlate motion of s-plane zeros with the motion of their z-plane counterparts. Therefore, in the following text, poles and zeros will be pictured on the s-plane, each having its own s-plane position, but the zeros due to double steps will be labeled "z-plane zeros" to denote their multiplicity, and for clarity the poles will be labeled "s-plane poles." Also, to give immediate meaning to the distances involved between zeros, in the case where $\beta=\pm n\pi/T$, the $n=1$ zeros will also be shown. All zeros are equally spaced.

This notation is shown in Fig. 1. Fig. 1 gives the pole-zero representations of various combinations of steps in the time domain.

Mathematically, it has been shown that zeros can be generated. The thesis herein is that these zeros can be used to determine pole locations of a system by superimposing them on the system poles and then noting their location to obtain the corresponding pole locations.

As a principle it can be stated that

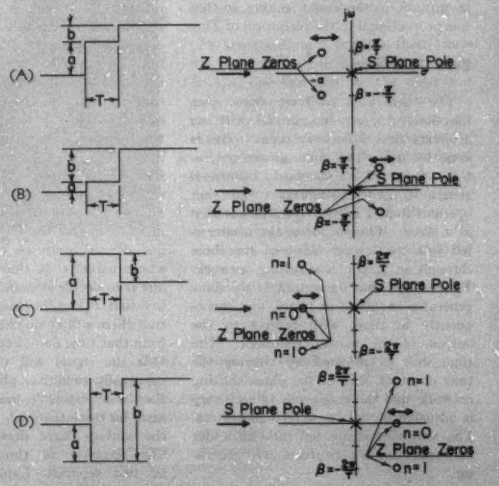


Fig. 1. s-plane pole-zero representations of double steps in the time domain. "z-plane zeros" denotes an infinite column of zeros, each separated from the next by $2\pi/T$

A—Minimum-phase complex zeros

B—Nonminimum-phase complex zeros

C—Minimum-phase real zero

D—Nonminimum-phase real zero

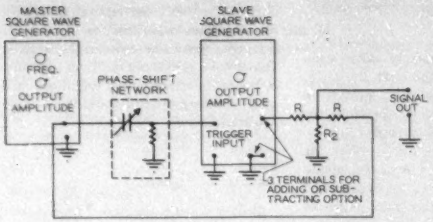


Fig. 2 (above). Zero-generator

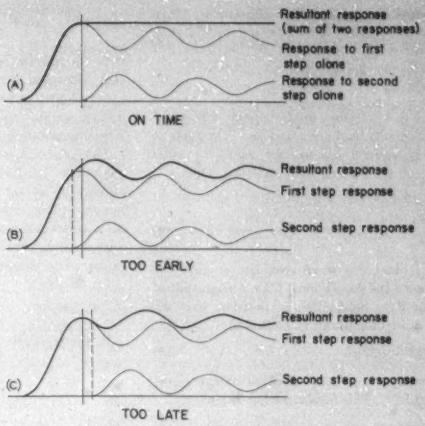


Fig. 3 (right). Summation of step responses

after the last step input (thus the zeros) has been added, if every s-plane pole has a superimposed s-plane zero, the output transient will have reached its final value tangentially, and will remain constant thereafter. This is called dead-beat response; i.e., all derivatives of the output simultaneously become identically equal to zero. Using this principle, the null condition (i.e., when the zeros are exactly on top of the poles) can be determined. In this case the null is detected visually.

To adjust the zeros until they are superimposed on top of the poles, the double-pulse spacing and amplitude ratio can be varied. If the timing is held constant and only the amplitude ratio is varied, the β value for the zeros will remain constant, and they will move on a locus of variable α which is parallel to the σ axis.

If the amplitude ratio is held constant and only the delay of the time of the second pulse is varied, the zeros will move on an hyperbola in the s-plane, asymptotic to the σ and $j\omega$ axes, so that α is proportional to the reciprocal of T.

Zero Generator

The generator delivers two step functions of given magnitudes with an arbitrary time T between them. This is done by using two square-wave generators. One is a standard calibrated generator, used as a master. The other operates from a trigger signal and is used as a slave. The output of the master is fed into the trigger input of the slave through an R/C phase-shifting network. The result is that the output of the slave generator is shifted in phase, and consequently in time, with respect to the output of the master generator. The time shift is adjusted by varying the time constant R/C of the phase-shifting network, and the magnitude of each step is adjusted at the respective generators. The two signals are fed into an adder network; this constitutes a zero generator."

Fig. 2 shows the wiring diagram of this system.

Application

As examples of the technique, tests on two systems are presented here. One system has one complex pole pair, and the other has a complex pole pair plus one real pole.

The method of investigation is to observe on an oscilloscope the response of the system to the zero-generator signal. The system acts as though a step is first introduced, and then suddenly at time T the zeros are introduced. A problem arises concerning the interpretation of the function given by $[aU(O)+bU(T)]$. In the time domain this is a step of magnitude a at time zero to which is added a step of magnitude b at time T. What about the Laplace transform? Can one say it is a/s for $0\leq t\leq T$ and $1/s\ (a+be^{-sT})$ for $t\geq T$ [i.e., a pole at the origin for time $O\leq t\leq T$ and the zeros suddenly appear at $t=T$]? Strictly speaking, no. One must recall that the Laplace transform is not defined until $t=\infty$ [including the whole universe of time]. Therefore, if one is to decide at some time $O\leq t\leq T$ to say that $F(s)=a/s$, then he is assuming that there will be no change in the signal from that time on. Yet here it is known that the signal will change and thus eventually yield the s-plane zeros. What then is a justifiable mathematical argument for this situation? For convenience the authors have decided to consider this response as though it operates in two separate Laplace "universes" [i.e., including all time >0]: universe no. 1 with a pole at the origin plus the system poles, universe no. 2 with a pole at the origin, the added zeros, plus the system poles. Then it is viewed as though the response is taking place in universe no. 1 for $0\leq t\leq T$ and in universe no. 2 for $t>T$. If the zeros in the signal exactly cancel the system poles, the system loses its characteristics associated with the poles, and immediately has a dead-beat response. This is demonstrated in the following experimental results.

Analysis of Step Response

If the response to the zero-generator signal is thought of as the addition of the response to the first step and the response to the second step, then the following will be noted:

If the second step occurs at exactly the correct time (i.e., to excite oscillations 180 degrees out of phase) and with the correct magnitude, then the oscillation components will cancel and there will be a dead-beat response; see Fig. 3(A). If this second step were too large in magnitude, cancellation would not occur, due to the magnitude differences, and a negative cosine wave oscillation would result, counting time from the instant of the second step. If the second step were too small in magnitude, again cancellation would not occur, and a positive cosine wave oscillation would result.

In the s-plane this means that the β of the zeros is correct, but the zeros are

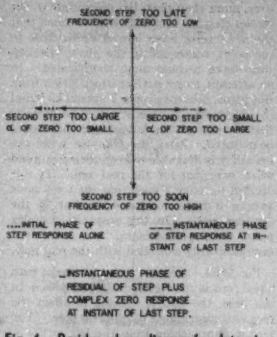

SECOND STEP TOO LATE
FREQUENCY OF ZERO TOO LOW
SECOND STEP TOO LARGE
α OF ZERO TOO SMALL
SECOND STEP TOO SMALL
α OF ZERO TOO LARGE
SECOND STEP TOO SOON
FREQUENCY OF ZERO TOO HIGH
.... INITIAL PHASE OF STEP RESPONSE ALONE
___ INSTANTANEOUS PHASE OF STEP RESPONSE AT INSTANT OF LAST STEP
_ INSTANTANEOUS PHASE OF RESIDUAL OF STEP PLUS COMPLEX ZERO RESPONSE AT INSTANT OF LAST STEP.

Fig. 4. Residue phase diagram for determining corrective procedure

respectively to the right of the poles (180 degree phase shift) and then to the left of the poles (no phase shift).

If the second step is too early in time then one notices a change in the response prior to the time the first transient reaches its peak value. Now, due to phase differences the two do not cancel each other, and a positive sine-wave oscillation results, again counting time from the instant of the second step; see Fig. 3(B). The resultant lags the first step response oscillation by approximately 90 degrees in phase, hence the transition from positive cosine to positive sine, because the zero is almost directly above the pole in the s-plane.

If the second step is too late in time then one notices a change in the response after the first transients' peak time. Again the two do not cancel due to phase differences and a negative sine-wave oscillation results; see Fig. 3(C). The resultant leads the original oscillation by approximately 90 degrees in phase, hence the transition from positive cosine to negative sine, because the zero is essentially below the pole in the s-plane.

The magnitude of the second step has its corresponding effect on the second and third cases.

A rule of thumb, then, would be first of all to have the steps occur with $T = t_{peak}$, then the proper b/a could be found to give the required dead-beat response; see Appendix II.

For these simple cases, one can construct a phasor diagram upon which the phase of the residual at the instant of the second pulse could be plotted. Consider Fig. 4, with the axes marked too early, too small, too late, and too large, as shown. If the phase of the residual at the instant of the last step were shown on this diagram, then the necessary corrective action for cancellation could be determined. This procedure will be illustrated in the following examples. By convention, the positive cosine wave will be used as the reference for measuring phase.

Experiment 1: System with One Complex Pole Pair

First the response to a step input is obtained. The response is studied and in this case it is noted that it is characteristic of a second order-system, i.e., a complex set of poles. Therefore, it is decided to sweep the range with complex zeros. An approximate region for the poles may be determined by finding the oscillation frequency and the characteristic damping from the step response. The step responses of second-order systems is given in Appendix I. The range of frequency and damping is swept until a null condition is observed, which determines the pole locations.

The experiment is described in Fig. 5:

Fig. 5(A). Step response of the system. Roughly estimate resonant frequency ω_n and damping factor ζ from Fig. 5(A) and thus derive approximate location for the poles.

Fig. 5(B). Introduce s-plane zeros in this region and note their effect. Proximity of the zeros to the poles is noted by their effect on the response. In this case condition 3(B) is noted; i.e., the second step is too early in time. Considering the phasor diagram approach, note that the phase of the resultant response at the instant of the second step is a little less than 180 degrees lag. This corresponds to the third quadrant of the phasor diagram. Therefore, the time T should be increased a little, and the magnitude of the second step should be decreased somewhat in an attempt to obtain perfect cancellation. This was done.

Fig. 5(C). Dead-beat response results. One can determine the pole locations from the zero-generator dials and from the relationship $\alpha = 1/T \ln b/a$ and $\beta = (2n+1)\pi/T$.

At this point the question arises, what value is given to n? One should note that dead-beat response occurs after one-half cycle in Fig. 5(C), and no oscillations are present. This corresponds to an n of zero.

When the time T is made three times as great, with a corresponding reduction in b/a, one can observe the dead-beat response shown in Fig. 5(D). Here there are 1 1/2 cycles of oscillation, an addition of 2π radians. This corresponds to an n of unity. Although this produced dead-beat response, it does not do so in the minimum possible time. The value of n is the number of extra periods (cycles) in the transient response above the minimum possible time for dead-beat response.

In Fig. 5(E), one may notice that two

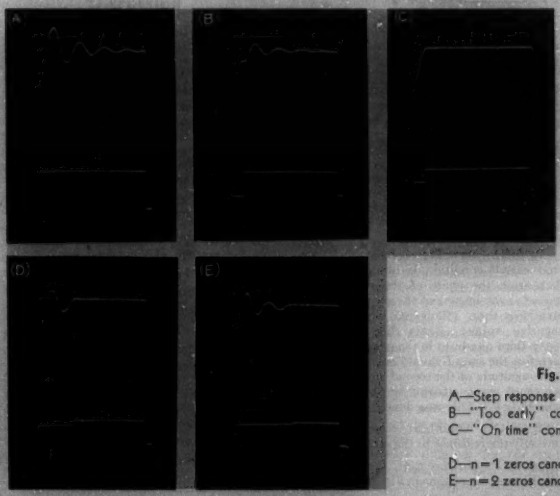


Fig. 5. System with complex pole pair

A—Step response
B—"Too early" condition (complex zeros to left of poles)
C—"On time" condition (complex zeros exactly cancel complex poles)
D—$n = 1$ zeros cancel complex poles
E—$n = 2$ zeros cancel complex poles

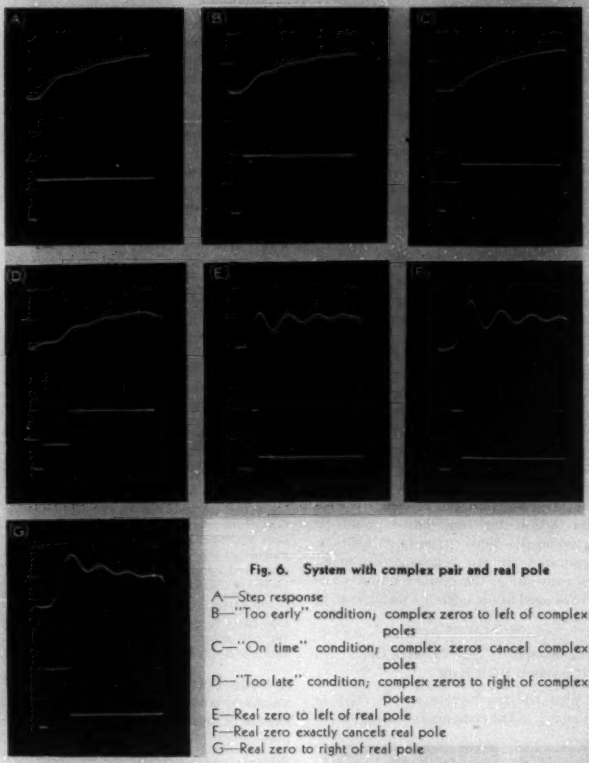


Fig. 6. System with complex pair and real pole

A—Step response
B—"Too early" condition; complex zeros to left of complex poles
C—"On time" condition; complex zeros cancel complex poles
D—"Too late" condition; complex zeros to right of complex poles
E—Real zero to left of real pole
F—Real zero exactly cancels real pole
G—Real zero to right of real pole

extra cycles are completed before dead-beat response; this is an addition of 4π radians, or $n=2$.

The pole locations can be determined directly from any of these conditions. However, the experimenter should, upon seeing condition 5(D) or 5(E) shorten the time T and push the zeros up in the s-plane until he obtains condition 5(C), for ease in computation.

Experiment 2: System with One Complex Pole Pair, and One Real Pole

First the response to a step input is obtained. The response is studied, and in this case it is noted that it is characteristic of a system containing a complex pair of poles and a real pole. It is further noted that the α of the real pole is of the same order as the α of the complex poles since both the transients have nearly the same decay time. Since the slope never goes negative, the α of the real pole is probably less than the α of the complex poles. It is decided to sweep the range with complex zeros to find the complex poles, then with a real zero to find the real pole.

The experiment is described in Fig. 6:

Fig. 6(A). Step response of the system. Proceed as before to locate the complex poles.

Fig. 6(B). As in Figs. 3(B) and 5(B), the "too early" condition is observed. Note that again the phase of the resultant response at the instant of the second step is a little less than 180-degree lag (almost a negative cosine). Again this corresponds to the third quadrant in the phasor diagram, therefore the time T should be increased and the magnitude of the second step should be decreased in an attempt to obtain perfect cancellation. This was done.

Fig. 6(C). Dead-beat response results and only the response to the real pole is left.

Fig. 6(D). In this case, the "too late" condition is observed as in Fig. 3(C). However, more than that, the magnitude of the second step has been increased so much that the $n=1$ zeros have moved down close to the poles, consequently the phase relationships have become more complicated. In an attempt to get perfect cancellation from this point, first the proper T should be found (by decreasing t), then it would be obvious that the second step is too large and should be reduced. After the complex poles are located it is desirable to keep their β in mind when sweeping for the real zero. It was decided to make T such that the $n=1$ zeros have a β large compared to that of the complex poles. In this way the complex poles were predominant, and were not simultaneously cancelled with the real pole.

Fig. 6(E). Now the system is swept with a real zero in the region of the real pole. In this case, the zero is starting to affect the response contributed by the real pole. The zero is approaching from the left in the s-plane. This is noted by the increased rise time and increase in oscillation amplitude.

Fig. 6(F). The real pole is exactly cancelled and only the contribution of the complex poles remains. Note that the center line of the oscillation is horizontal.

Fig. 6(G). The real zero has passed the real pole and is in near the origin of the s-plane. This fact is borne out by the sharp rise time and the negative slope of the center line of the oscillation. The effect of the real zero dominates that of the real pole.

Conclusions

A technique using a signal to determine directly the system characteristic pole locations has been presented. This technique offers a means to by-pass the tedious process of taking steady-state frequency response and/or transient response data, synthesizing equations, and factoring these equations when trying to find an analytical description of a system. The complex-zero generator permits high-accuracy measurements in systems with low Q's (0.7 to 5). It is like a Q-meter except that in this case the accuracy of the Q measurement increases as the Q decreases.

The method was here applied to relatively simple systems, however the authors feel that its extension to more complex systems will follow after further techniques and insight are developed. The possibilities for this method seem quite promising. Perhaps a null method other than visual can be developed, hence leading to an automatic scanning type of instrument. The advantages following this achievement are self-evident.

The next step in this endeavor should be to find a similar way to determine system zero locations. By analogy one would expect to superimpose signal poles on system zeros to determine the latters' position, however, the problem here is generating complex poles. This is not

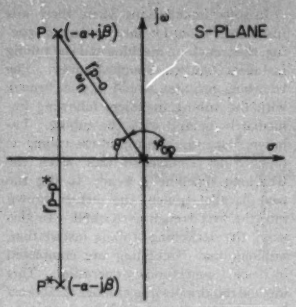


Fig. 7. s-plane nomenclature

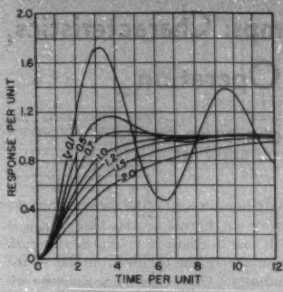


Fig. 8. Response to step input of complex pole pair with $0.1 \leq \zeta \leq 2$

easy. Nevertheless, once all of the pole locations are found using the complex-zero generator, the zero locations may be found rather easily from frequency response data.

Appendix I

"Characteristic damping," or "dimensionless damping factor" for a second-order underdamped system is defined as:

$$\zeta = \cos\theta = \frac{\alpha}{\omega_n} \tag{21}$$

θ is the angle between the negative real axis and the line connecting the origin and one of the complex poles as shown in Fig. 7.

Step responses for typical ζ's are given in Fig. 8.

Appendix II

Since β is given by $\beta = \pi/T$, this condition implies that $\beta = \pi/t_{peak}$. This relationship is indeed true and can be found in the following manner.

Given a system that can be represented in the s-plane by a complex pole pair in the left half plane, the response to a unit step input can be expressed in terms of s-plane angles and distances as follows (see Fig. 7):

$$C(t) = 1 + \frac{2r_{P-0}}{r_{P-P^*}} e^{-\alpha t} \sin(\beta T - \phi_{0P}) \tag{22}$$

This form can be derived from the standard inversion formulas by making substitutions based on the geometry of the s-plane configuration. Differentiating in the time domain is like multiplying by s in the Laplace domain. Therefore, the pole at the origin introduced by the signal is cancelled by differentiating, and $C'(t)$ obtained directly from the resulting s-plane pattern as:

$$C(t) = \frac{2r_{P-0}^2}{r_{P-P^*}} e^{-\alpha t} \sin(\beta T) \tag{23}$$

This expression goes to zero when $\sin \beta T$ does. This occurs when $\beta T = n\pi$. Here the first time larger than zero is desired, therefore set $t = \pi/\beta$ and $t_{peak} = \pi/\beta$. (The method here has been to set the derivative to zero for the maximum point.)

References

1. Response of Physical Systems (book), J. D. Trimmer. John Wiley & Sons, Inc., New York, N. Y., 1950.

2. Control System Dynamics (book), Walter R. Evans. McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

3. Automatic Feedback Control System Synthesis (book), John G. Truxal. McGraw-Hill Book Company, Inc., 1955.

4. Direct Synthesis Through Block Diagram Substitutions, O. J. M. Smith. *IRE Convention Record*, Institute of Radio Engineers, New York, N. Y., 1957.

5. Mixed Distributed and Lumped Parameter Systems, O. J. M. Smith. *IRE WESCON Convention Record*, Institute of Radio Engineers, 1957.

6. Poincast Control of Damped Oscillatory Systems, O. J. M. Smith. *IRE Proceedings*, Institute of Radio Engineers, vol. 45, no. 9, Sept. 1957.

7. Analogue Study of Dead-Beat Poincast Control, O. J. M. Smith, George Tallman. *IRE Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, Spring 1958

Discussion

Keisuke Izawa (Purdue University, Lafayette, Ind.): The authors have performed a valuable service in determining pole locations of system transfer functions. The procedure must also be applicable to the investigation of the predominant modes of the system response unaffected by the dead time of the system. This approach will be useful in determining the system parameters when the functional form of the system transfer functions are known.

I feel that to select the three parameters, a, b, and T of the complex-zero signal, is redundant but that two parameters, for example, b/a and T, are sufficient. It would be interesting to have the explanation on how the method works for testing (1) two or more systems with same poles but different zeros and (2) systems with multiple poles.

G. G. Lendaris and O. J. M. Smith: We would like to thank Mr. Izawa for his comment concerning the advantage of this method in systems with dead time. He has called attention to an important application of this method, pointing out that it is not necessary to separate the excessive phase lag of the dead time from the phase lag of the lumped-parameter poles in order to measure the time constants and resonant frequencies of the lumped-parameter poles.

He also mentioned that this approach would be useful in determining the system parameters when the functional form of the system transfer function is known. This is indeed an important observation. Using this technique one has only to design his system, build it with adjustable components in the desired region, set the signal generator to the desired pole location, and then observe the response of the system to this signal. The parameters are then adjusted until the poles are exactly cancelled by the zeros. In this way one can very quickly and accurately obtain his desired system. It has been tried in our laboratory and proved to be very useful indeed.

Although we showed the parameters a and b separately in the paper, we always used the ratio b/a as the only significant parameter. It is not important what the absolute magnitude of the steps is, so long as the system remains within its linear range of operation.

In comparing two systems with the same poles but different zeros, the shape of the step transient, and the shape of the transient with complex zeros that are not properly adjusted, are each different for the two cases. However, for each case, when the complex zeros exactly coincide with a complex pole-pair of a system, an oscillatory component in the transient disappears, and the remainder is a measure of the location of the system zeros and the remaining poles. It is therefore possible to measure the system pole positions by visual observation of the system response, in the presence of zeros.

It is not possible to measure directly the positions of the zeros of the system. However, in many cases it is possible to make a second set of measurements which determine the inverse transfer function, and during these measurements all of the system zeros can be located. As an example, if one drives a 2-terminal network with the voltage from a complex zero generator and observes the current into the two terminals, he can measure the poles of the admittance function. Conversely, if he drives the two terminals through a high impedance from the complex zero generator, and measures the voltage across the terminals, he can determine the poles of the impedance function.

We are presently investigating the use of this method on systems with second-order poles on the real axis, with second-order complex pole pairs, and with two pairs of complex poles with arbitrary orientations with respect to each other. The results are quite pleasing and it is planned to present them in another paper in the near future.

Earth Source Heat Pumps: Characteristics, Design, and Operation

E. A. FREUND
NONMEMBER AIEE

G. S. WHITLOW
FELLOW AIEE

TWENTY YEARS AGO the programs of many engineering meetings included papers on air conditioning. Air conditioning was "billed" as the business which would pull the country out of the depression. Today the heat pump appears with almost equal frequency in the discussions of technical societies. The claims made for it are more realistic than were those for air conditioning.

The heat pump will have a greater impact on the lives of the customers of the utility and on the utility than air conditioning. This is a broad statement, but the experience with heat pumps on the Union Electric system, even though that experience is still limited in scope, tends to bear out the statement.

The heat pump as it is known today is only in the early stages of its technical development. A coefficient of performance of 3 is talked of. It will surely be much higher before the heat pump has reached the present age of air conditioning. This fact assures that the operating costs of heat pumps will decline in the future. Industry must strive diligently to reduce the cost of operating electric appliances; only by that means can a reasonable degree of saturation of such appliances be achieved.

The heat pump is attractive to the utility because in most areas of the country it is a year-round load. The authors believe that in this area the diversified annual load factor of the heat pump in residential service may well reach 30%. A saturation of 16% could double the present national average annual consumption of 2,500 kw-hr (kilowatt-hours) per residential customer.

These observations apply to heat pumps generally, regardless of the heat source. The following discussion deals mainly with heat pumps using earth as a source of heat. The authors' company is not wholly convinced that earth is the one and only heat source for the territory but the work was started on this basis and it is felt that to continue to explore this path would serve the common interest of the industry. Others are doing similar work as well as exploring the air and water heat sources.

There are two general types of ground coil applications. One is generally termed a "direct-expansion" application because the refrigerant is circulated through pipes buried in the ground. The other is generally termed "liquid circulation" because a nonfreezing liquid is circulated through the ground coils. This company has had experience with both. Each type has it advantages and its problems.

Many improvements have been made in methods of ground coil installation since the first was installed. Undoubtedly, further improvements will be made when experienced builders and contractors have had the opportunity to install substantial numbers of ground coils and to gain experience in this field.

Installation Methods

To illustrate the installation methods this company has used, refer to the schematic ground coil diagram shown in Fig. 1, and the plot plans of two actual installations as shown in Figs. 2 and 3.

The best method of trenching seems to be with a Cleveland-type wheel trencher or with a revolving chain digger on a boom mounted on the rear of a tractor. One example of the latter type is known as a "trench-hog." Both of these machines are capable of digging a 15-inch trench to a depth of 4 or 5 feet. Another type of machine known as a "back-hole digger," also mounted on the rear of a tractor, has been used occasionally and does a satisfactory job, although requiring a much longer time.

If the trenches are laid out on 8-foot or 10-foot centers, the islands between the trenches should be capable of holding the dirt removed from the trenches. This method of course, simplifies the ground coil installation since it permits the trenching machine to finish its work before the actual installation of tubing.

In some installations where there was insufficient room for this amount of spacing, the trenching machine and the tubing installers worked simultaneously. The trenching machine would dig one trench with the tubing installers following immediately behind with the tubing. Laborers place sand on top of the tubing as quickly as possible so that when the trenching machine is ready to dig the next parallel trench, the dirt is thrown into the first trench as backfill. In this way, the trenching, tubing installation, sanding, and backfilling are completed in one continuous operation. This method requires a large number of laborers for smoothing and cleaning out the trenches and carrying sand so that the trenching machine may continue at a normal pace and is not unnecessarily delayed.

The tubing is made into the required lengths in advance of the trenching operation. This prefabrication is done by an experienced refrigeration service organization. Hard solder and sweat couplings are used. The tubing is left coiled so that when the soldering operation is completed, one long coil results. This coil is then tested for leaks by applying a pressure of 250 psi (pounds per square inch) of dry nitrogen for several days. The joints are also tested with suds to be doubly certain that no leaks exist. At the end of the test the ends of the tubing are sealed with pressure remaining in the tubing. The coils are then ready to be unrolled in the trenches.

The actual unrolling of the tubing is usually done by a refrigeration man or plumber with the help of one or more laborers. One end is unrolled for several feet and this end is extended through the hole in the basement wall. The large coil is then rolled along the bottom of the trench with the proper care taken to avoid kinks. The other end of the coil is also

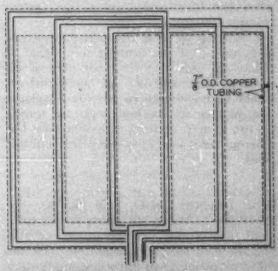


Fig. 1. Typical ground coil for 5-hp compressor

Paper 58-1010, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Sacramento, Calif., August 19-22, 1958. Manuscript submitted Sept. 30, 1955; made available for printing June 18, 1958.

E. A. FREUND and G. S. WHITLOW are with the Union Electric Company of Missouri, St. Louis, Mo.

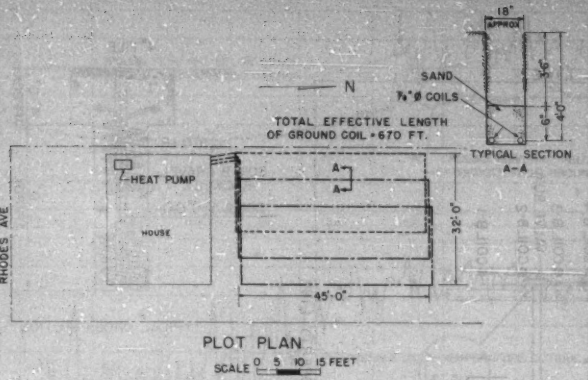


Fig. 2. Ground coil plan, at 5871 Rhodes Ave., St. Louis, Mo., Installation C

extended into the basement. The tubing is usually left in this state inside the basement until the manifolds are to be installed, at which time the tubing is cut off evenly about 8 to 12 inches from the wall.

Several methods have been used for passage of the tubing through the basement wall. One method is to fabricate a rectangular opening in the wall before it is poured. When the tubing has all been extended through this opening and arranged properly, the hole is filled with expanding concrete. Another method is to drill individual holes through the wall, extend the tubing through these holes, and seal each hole carefully. Perhaps the best method was used on a recent installation. Before the basement wall was poured, short lengths of $1^1/_4$-inch OD copper pipe were placed through the form at each spot where a ground coil line was to enter the basement. The 7/8-inch OD ground coil tubes were then passed through these pipes and soft solder or waterproofing used as a seal.

An initial covering of 8 to 10 inches of sand is placed over the tubing in the trench, and the backfill completed with the orginal dirt removed. The use of sand assures complete contact around the tubing and aids in moisture retention. If the trenches are carefully watersettled after backfilling, little future settling will occur.

Complete information on the installation cost of ground coils is unavailable. The best information now available indicates that ground coils can be installed for approximately $0.75 to $0.80 per foot of tubing or for $0.90 to $1.00 per effective foot of ground coil. The term effective length was originated in order best to express the sizes of various ground coils. The effective length of a particular ground coil is merely twice the total length of trenches, including run-outs to the house. Thus, each foot of trench where there are more than two tubes, such as run-out or header trenches, is assumed to be only as effective in transferring heat as a foot of longitudinal trench containing only the two tubes. This assumption is based on the theory that the limiting factor in the amount of heat transfer is the ability of the surrounding earth, rather than the amount of tubing, to conduct heat.

A ground coil for a typical installation where two 5-hp (horsepower) compressors are to be used would have an effective length of about 2,000 feet. Thus, its cost would be about $2,000. This total cost would be composed of the following approximate units:

Trenching and backfilling	$ 520
Copper tubing	860
Prefabrication and testing of coil	80
Sand	100
Hand labor	240
Installation labor	200
	$2,000

This cost of course, is based on present costs of materials and on labor rates prevailing in the St. Louis area. Better installation techniques could probably decrease all costs except that of the copper tubing. This cost is believed to be no greater than comparable costs for air coils, defrosting controls, fan, and insulated duct.

Types of Ground Coils

The first experimental ground coil installed consisted of 10 helical coils of $1^1/_4$-inch steel pipe; see Fig. 3(A). Each coil was about 14 inches in diameter and 12 feet long. These were installed vertically about 10 feet apart so that the tops were 4 feet below the surface of the ground. A nonfreezing liquid was circulated in two parallel circuits of five coils each. The operation of the ground coil was disappointing because it soon became evident that during winter operation the coil could not extract an adequate amount of heat from the earth without reducing refrigerant suction temperatures below economical limits.

The second ground coil is a horizontal ground coil of entirely different design; see Fig. 3(B). It consists of about 1,600 feet of $1^1/_4$-inch OD copper pipe in four sections of 400 feet each. The pipes were laid in trenches 4 feet deep and 1 foot wide. The selected depth of 4 feet was based on several considerations. It had been deduced and proven that in cold weather heat would not flow into the ground area fast enough to maintain earth temperatures above freezing. It was reasoned, then, that depths much below the maximum St. Louis frost line of 3 feet could not be justified. Although greater depths would give higher earth ambients and theoretically higher performance factors, recovery would be more rapid with coils closer to the surface. Also, 4 feet is a depth within easy limits of standard trenching machines. The first 10 or 12 inches of backfill was sand in order to get more intimate contact and better moisture retention.

A number of thermocouples were installed with this second ground coil. At two places in the ground coil area, thermocouples were placed 4, 6, 8, and 12 feet below the surface. Those 4 feet below grade were actually in the bottom of the trenches. In addition, thermocouples were installed remote from the ground coil at the same depths in order to establish reference ambient temperatures. The readings of these thermocouples have been continuously recorded since 1948.

A graph is included which shows the earth temperatures adjacent to the ground coil, the corresponding ambient earth temperatures and the average daily air temperatures recorded during the year 1950; see Fig. 4. These are believed to be typical for ground coils in the St. Louis area.

It was obvious that savings in both first cost and operating cost could be realized if the refrigerant were circulated directly

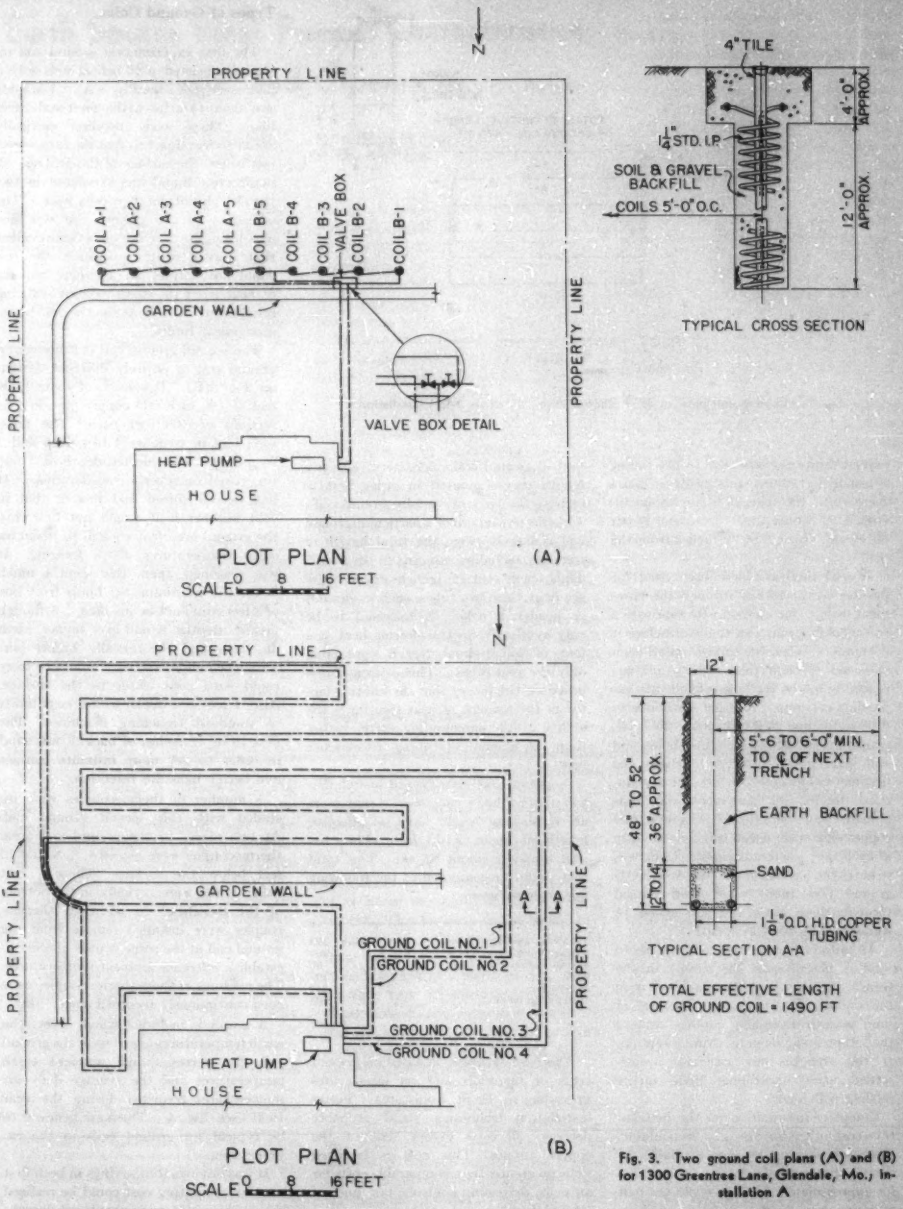


Fig. 3. Two ground coil plans (A) and (B) for 1300 Greentree Lane, Glendale, Mo., installation A

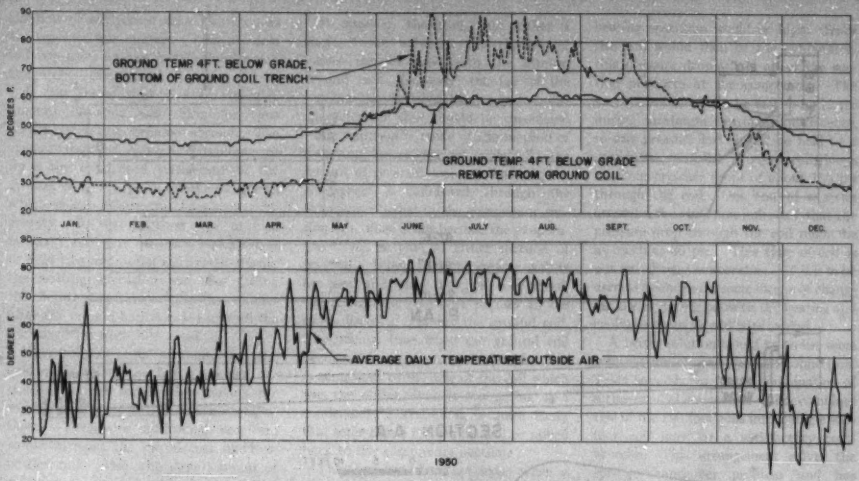

GROUND TEMP 4FT. BELOW GRADE, BOTTOM OF GROUND COIL TRENCH
GROUND TEMP 4FT. BELOW GRADE REMOTE FROM GROUND COIL
DEGREES F.
JAN. FEB. MAR. APR. MAY JUNE JULY AUG. SEPT. OCT. NOV. DEC.
AVERAGE DAILY TEMPERATURE-OUTSIDE AIR
1950

Fig. 4. Ground and outside air temperature, at 1300 Greentree Lane, Glendale, Mo.; installation A

in the ground coil. A large water-to-refrigerant heat exchanger and a circulating pump and motor could be eliminated if the antifreeze solution were not to be used. The operating costs would be less because of the elimination of the pump motor, and also because the compressor in a direct-expansion system would operate at higher suction pressures since there would be no intermediate heat transfer medium.

Little data were available regarding the problems which might be encountered when refrigerant was introduced directly into the ground coil. One of the major concerns was that of oil return. It was feared that entrained oil might become trapped in the ground coil, reducing its heat transfer capacity and leaving the compressor crankcase short of oil. It was decided to design the ground coil for velocities sufficiently high to assure the return of this oil, but at the same time not so high as to cause excessive pressure drops.

Also, there was fear of leaks. However, it was assumed that if the joints were carefully made with the proper solder and tested, no trouble of this kind should occur. Experience has borne this out.

The first experimental direct-expansion ground coil was installed and connected to a 3-hp compressor at a small heat pump laboratory set up by Union Electric engineers. This coil consisted of 900 feet of 7/8-inch OD copper tubing installed in 450 feet of trench. There were six separate loops, each about 150 feet long, and terminating at two common headers. Liquid refrigerant was introduced to the coil through a multiport expansion valve, with the compressor suction connected to the other end. Each of the six loops was separately valved so that the amount of ground coil could be varied.

After about 5 months of operation with the complete coil in service, it became apparent that the 900-foot length was more than sufficient for a 3-hp compressor. Half this amount was tried, but it was quickly determined that this was too small to provide sufficient evaporating surface. The next trial was with two thirds of the entire coil, or 600 feet. This amount of ground coil appeared to balance the compressor capacity for the heating cycle. To determine if the heat in the ground could be depleted by long hours of operation, an extremely severe test was planned with the use of the 600 feet of ground coil. The compressor was allowed to operate continuously on the heating cycle for approximately 16,000 hours. Pressure recorders showed that the suction would rise each fall, then slowly diminish during the winter, reaching a minimum value of about 10 psi in the spring.

After this test had been in progress about 3 months, one of the trenches was opened to determine the condition of the soil surrounding the two tubes in the trench. An elliptical mass of frozen earth with axes of 28 inches horizontally and 21 inches vertically was found.

The ground coil design for several subsequent residential heat pump installations was based upon the satisfactory operation of this laboratory test. The first of these went into service in the fall of 1950 in a small residence in St. Louis. This installation consisted of a custom-built package unit with two 3-hp compressors connected to 900 feet of 7/8-inch OD copper tubing installed in a 30- by 50-foot area.

Two more experimental ground coils were installed at the heat pump laboratory in the winter of 1952-1953. The first of these was identical to the original ground coil at this location, except that it was buried $2^1/_2$ feet, instead of 4 feet as was the original. It was connected to the same heat pump equipment so that its operation could be compared with the performance of the original coil. Some heating operation data have been obtained, but outdoor temperatures during this period were unusually high and no definite conclusions have been reached. Further tests are in progress.

The third ground coil at the laboratory consists of three loops of 7/8-inch OD copper tubing, each approximately 200 feet long; see Fig. 5. Three 6-inch-wide trenches were dug away from the building and one loop of tubing place in each trench in such a way that the portion of

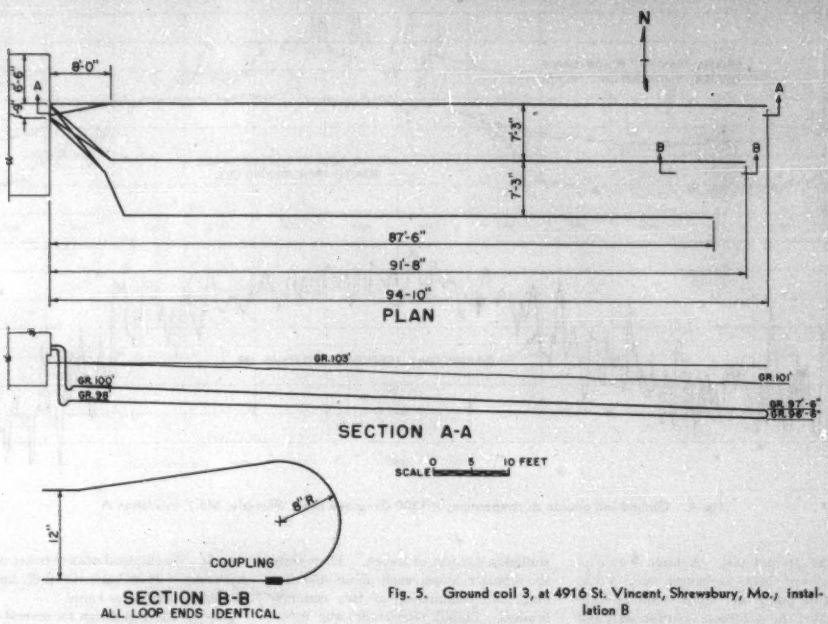


Fig. 5. Ground coil 3, at 4916 St. Vincent, Shrewsbury, Mo., installation B

the loop running away from the building is directly above the portion of the loop returning to the building. Therefore each loop is similar to a hairpin turned on its edge. At the building, the top tube of the loop is 3 feet below the surface and the bottom tube is 5 feet below the surface. At the end of the trench away from the building the coil is 4 feet deep. Sight glasses were placed in both the inlet and outlet pipes of each ground coil loop to permit observation of the refrigerant flow through the ground coil. This coil was connected to a second heat pump system at the laboratory, which also was built around a 3-hp compressor. This unit has been operated on both heating and cooling, but again no definite results are available.

The experience gained from these experimental ground coils has been used in designing subsequent installations. A minimum requirement has been assumed of 200 feet of 7/8-inch OD copper tubing per nominal horsepower of compressor capacity, slightly varied to adjust to the lot area and to other physical conditions as they exist for each heat pump installation. This crude design criterion has worked reasonably well. It should be emphasized that there is need for a ground coil design formula based on better techniques than can be supplied at present. The authors are glad to say that, since 1948, a comprehensive study based on laboratory tests, which takes soil characteristics into account, has been sponsored by a sister utility.

By the spring of 1953 there were a number of ground coil heat pumps operating in the St. Louis area. Some made use of a single 4-way valve for changeover from heating to cooling operation with a resulting reversal of the direction of the refrigerant flow in the ground coil. Others used two changeover valves, either two 3-way valves or a 3-way and a 4-way valve, and in these units the refrigerant flow through the ground coil was in the same direction for both heating and cooling. Some of the ground coils were practically level, some sloped down from the house, others up from the house, and one was pitched throughout its entire length. This latter coil was connected to a unit with two changeover valves so that the refrigerant flowed downward through the ground coil during both the heating and the cooling operations. Thus, there were several different ground coil types and connections in use. All seemed to perform well during heating operation.

When cooling is first needed in the spring, the temperature of the earth surrounding the ground coil is comparatively cool, about 50 or 55 F (degrees Fahrenheit) at a 4-foot depth, late in May, in St. Louis. This temperature is not influenced to any great extent by the fact that the ground coil has been used for heating and has been operating as an evaporator. By the time cooling is required, usually sufficient time has elapsed since the end of the heating season, so that the ground temperature adjacent to the ground coil has reached the ambient earth temperature.

During summer operation with these low condensing temperatures, several of the heat pumps failed to operate properly. The suction pressure dropped to zero psi, the head pressure about 50 to 60 psi, and no flow was visible in the sight glass. It appeared that the refrigerant vapor was condensing almost immediately after entering the ground coil, and in those coils which did not have a continuous pitch the liquid refrigerant was remaining in the ground coil. Thus,

the flow of refrigerant ceased as soon as the receiver was empty. This problem did not exist on the installation having the pitched ground coil.

It was found that if a considerable amount of additional refrigerant was added, operation became normal. The system requires enough refrigerant to force the condensed refrigerant in the ground coil to a point in the coil from which it can return to the receiver by gravity. It was necessary to provide auxiliary receivers on some installations to hold this additional refrigerant during the heating operation and also during the cooling operation later in the summer after the ground had warmed up and the discharge pressure had risen.

This solution was adequate for the existing installations, but it seemed desirable to find some other means for solving the problem on future installations. Larger or additional receivers would increase the initial cost and size of the unit. Also, the large amount of refrigerant required adds to the initial cost and offers a potentially large replacement cost in case of leaks. The best solution for future installations, it was decided, was the pitched ground coil.

The question then arose as to the possibility of using only a 4-way valve for reversal with this coil, since this arrangement would result in upward flow of the refrigerant as it passed through the coil during one of the cycles. It was obvious that the refrigerant should flow downward as it condensed during the cooling cycle, so the question then became one of refrigerant flow upward, as it evaporated during the heating cycle. All authorities were agreed that this procedure should offer no difficulties. Several months' operation of a unit of this type showed satisfactory results.

It appears, therefore, that either a single 4-way valve or two changeover valves can be used with the pitched ground coil. However, the use of the two changeover valves has one undesirable characteristic that should be considered when selection is made of the method of reversing the refrigerant cycle. At the instant of reversal, the coil that was the condenser is connected through the changeover valve to the compressor suction line, since it has become the evaporator. This is true of either method of reversal. When two changeover valves are used and when the cycle is reversed, the refrigerant continues to flow in the same direction through the ground coil. This means that when the ground coil becomes the evaporator, the suction line is connected to the side of the coil which was the outlet when it was acting as a condenser. Therefore it is quite likely that some liquid refrigerant will be pulled back to the compressor suction.

This problem does not exist when a single 4-way valve is used. In this case the suction line connection is made to the coil which was the inlet when it was acting as a condenser, and in this portion of the coil the refrigerant is in the vapor state. The difference between the two methods of reversal is apparent from the schematic diagrams which are shown in Fig. 6.

Another method for eliminating this spring cooling problem was a suggested change in ground coil circuiting. Instead of using a number of parallel ground coil loops, one long loop would be used; in other words, the individual parallel loops would be connected in series. This could be done very easily at the headers in the basement or utility room. The disadvantage of this plan is that the pressure drop through the coil during heating operation would be high. Since this coil would then be the evaporator, this pressure drop would affect the suction pressures at the compressor. The suction pressure may be as low as 12 psi during maximum heating requirements, so any pressure loss should be held to a minimum. Parallel ground coil loops result in pressure drops of about $1^1/_2$ psi through the coil. One test of a series ground coil connection indicated that the pressure drop through the coil might be as much as 30 psi. This type of coil is not an efficient evaporator. If it is to be used as a condenser, some means of changing the circuiting between the heating and cooling seasons should be provided.

A better solution would be to use some combination of series-parallel connection. Tests have shown that a combination of series-parallel connection which combines two of the 200-foot coils into a single 400-foot coil may be a satisfactory compromise. This arrangement solves the spring changeover problem and has operated satisfactorily during past hot and dry summers. If the pressure drop through this coil during winter operation, when it is the evaporator, were found to be excessive, provision could be made for automatically operating valves to revert to the straight parallel coil. This has been successfully done on one installation. This solution could be applied not only to new installations but it would also be relatively simple to make the necessary changes on the ground coil headers of existing installations.

Another problem became apparent during the summers of 1953 and 1954, which in the St. Louis area were exceptionally hot and dry. A number of the heat pumps operated at head pressures considered higher than desirable. Discharge pressures of 180 psi were encountered

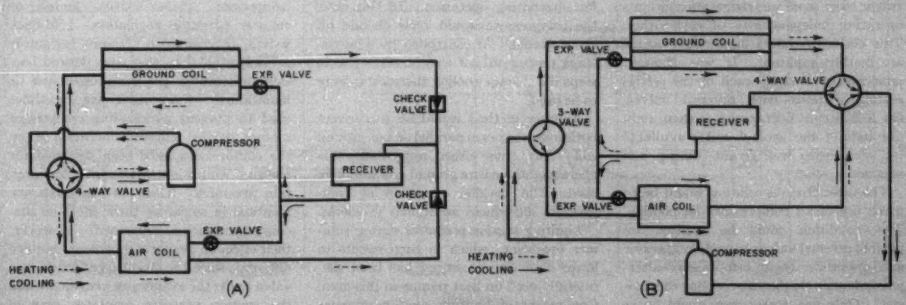


Fig. 6. Schematic diagrams of heat pump

A—Using only 4-way valve B—Using 4-way and 3-way valve

TABLE I

HEAT PUMP INSTALLATIONS IN THE ST. LOUIS AREA

SUMMARY OF ANNUAL ENERGY CONSUMPTION AND MAXIMUM DEMANDS

MAY 20, 1953 TO MAY 25, 1954

INSTALLATION	CONNECTED LOAD HP	HEAT LOSS (1) BTU/HR	DEMAND 15 MIN. MAX. KW	HEATING SEASON KWH (2)	COOLING SEASON KWH	ANNUAL KWH	ANNUAL COST (3)
E	10.75	85,110	12.72	15,310	6,780	22,090	387
K	6.50	34,750	7.00	7,460	1,750	9,210	161
I	6.75	68,780	7.60	6,980	3,240	10,220	179
F	6.25	69,260	6.40	9,340	670	10,010	175
C	6.50	40,000	4.30	5,390	4,850	10,240	179
J	10.75	79,010	10.08	16,230	5,600	21,830	382
G	21.50	157,030	18.24	25,340	14,420	39,760	696
D	12.50	80,580	12.50	21,170	7,910	29,080	509
H	16.25	150,610	19.80	27,330	8,760	36,090	632
AVERAGE	10.85	85,010	10.96	14,950	6,000	20,950	367

(1) CALCULATED IN ACCORDANCE WITH ASH & VE STANDARDS.

(2) ADJUSTED FOR A NORMAL HEATING SEASON OF 4584 DEGREE DAYS.

(3) ELECTRIC ENERGY AT 1.75¢ PER KWH.

NOTE: ENERGY AND DEMAND WERE CONTINUOUSLY RECORDED ON EACH INSTALLATION DURING ENTIRE PERIOD.

and on some units the pressure rose to 200 psi.

In the St. Louis area, residential winter heating requires approximately twice the amount of compressor capacity than is needed for summer air conditioning. Most of the installations are made up of two or more separate refrigeration circuits operating independently of each other. This condition offers advantages which are readily apparent. It was thought unnecessary to equip each of the refrigerating systems with reversal valves. On a 2-circuit heat pump system only one half of the ground coil is available for condensing refrigerant during the summer.

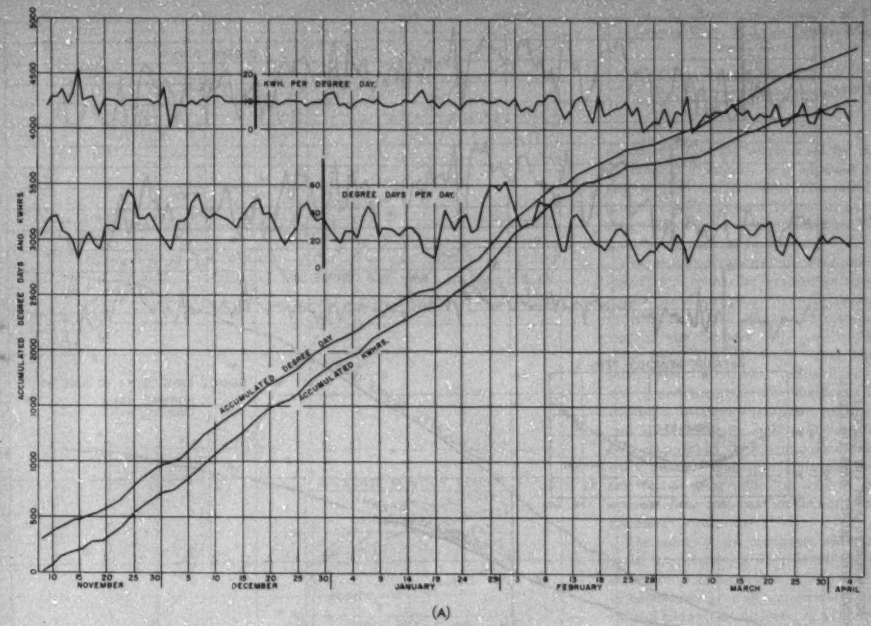
The most obvious solution would be to make use of all compressors for cooling. This condition could be realized by putting reversal valves on each compressor and operating these compressors alternately from a single-stage cooling thermostat. Only one compressor would be in operation at any time. A time clock could be used to alternate compressor operation. This method would double the available condensing surface with very little additional first cost merely by utilizing all the ground coil that is installed. A modification of this plan would be to equip both compressors with changeover valves but with no provision for alternating operation. In this case, the compressors would cycle on and off simultaneously, if controlled by a single-stage thermostat, or they would cycle in steps if a 2-stage cooling thermostat were to be used.

Another method would be to connect both compressors in parallel on one ground coil. Any time either compressor was operating, the entire ground coil would be used. The parallel operation of compressors introduces additional problems.

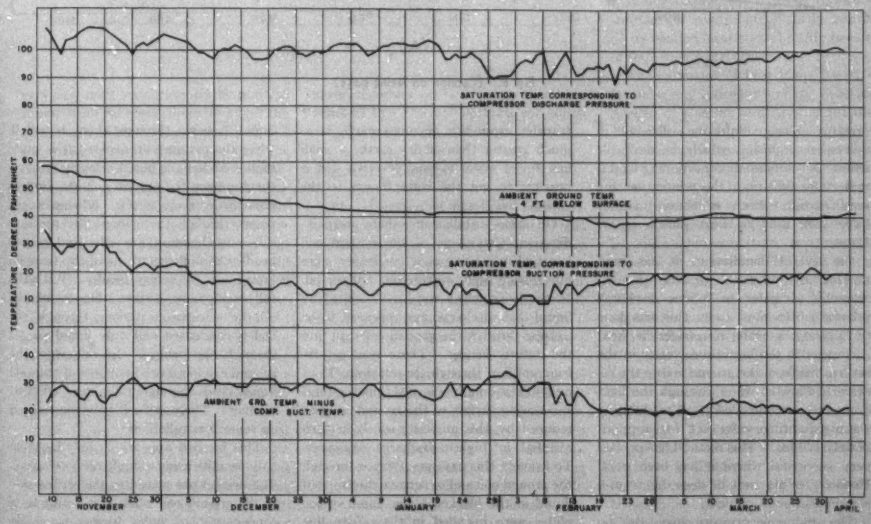
Limiting suction pressures during summer operation, which in turn results in lower discharge pressures, has been successfully used on heat pumps in this area. One means of limiting suction pressure is to use an expansion valve with a gas-charged power assembly having a maximum operating pressure of 25 or 30 psi. The suction pressure will be limited to this pressure and the discharge pressure will be much lower than if the unit were operating at 35 or 40 psi.

The other method is to use a pressure-limiting valve in the suction line at the compressor. These valves, known as suction pressure regulators, hold-back valves, or crankcase pressure regulators are adjustable, so that any desired suction pressure at the compressor can be maintained. These valves are sometimes used to prevent overloading compressor motors during summer operation when the compressors have been designed for efficient winter operation with low suction pressures. These two methods are identical in so far as their effect on discharge pressure is concerned. However, their effect on evaporator pressure is quite different, since the gas-charged expansion valve limits the evaporator pressure, while the suction pressure regulator permits normal evaporator pressures.

In either case, a lowering of the suction



(A)



(B)

Fig. 7. Record of heating operation during winter, 1950-1951, at 5871 Rhodes Ave., St. Louis, Mo., Installation C

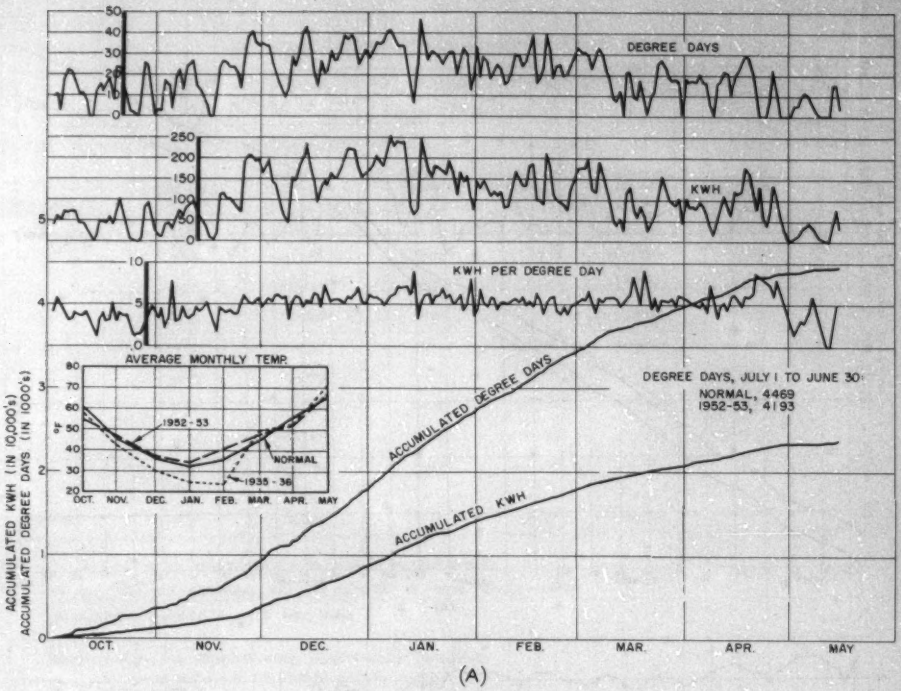


Fig. 8. (Caption on facing page)

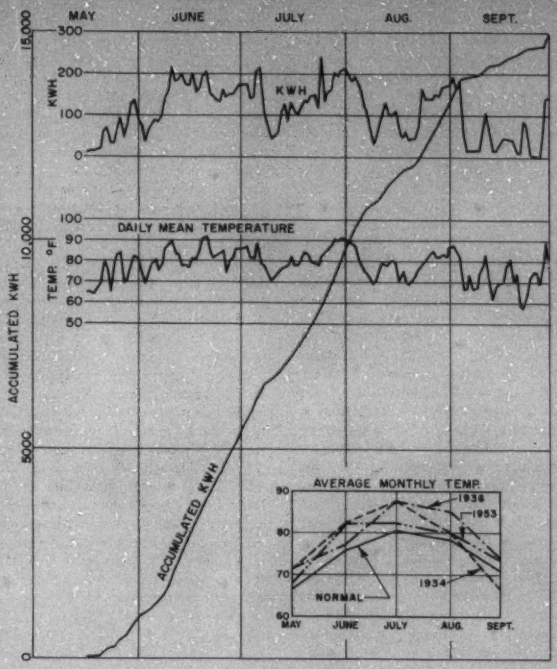
pressure at the compressor will result in decreased capacity which is desirable where the summer air-conditioning load is materially less than the winter heating requirements. Both of these methods have been used on heat pumps in St. Louis.

On several installations it has been found advantageous to use limited amounts of water to reduce discharge pressures. In most cases, this was done by placing a water-to-refrigerant heat exchanger in the liquid receiver or in the return line from the ground coil to the receiver. Flow of water through the heat exchanger is controlled by a modulating water-regulating valve set for approximately 170 psi. This method has proved very successful where it has been used. However, it may not be desirable from a commercial point of view.

The discharge pressure may also be reduced by running small quantities of water into the ground surrounding the ground coil tubing. Because the heat transfer capability of moist earth is so much greater than of dry earth, a small amount of water is very effective if it is possible to get the water down to the ground coil depth.

On some installations, there were indications that some of the ground coil loops when serving as a condenser were filling with liquid refrigerant and or oil. In the loops which were not blocked by liquid, hot discharge gas appeared to be passing through the ground coil and into the return header. There was no refrigerant flow through the remaining loops under these conditions, and thus the total condensing surface of the ground coil was reduced by this amount, which in turn resulted in higher discharge pressures. To prevent this passage of vapor through the ground coil and to remove the trapped liquid in the blocked loops, small orifice plates were installed in the return line of each ground coil loop immediately before the header. These orifices operate on the principle that the pressure drop of a liquid passing through them is small, while the pressure drop of a gas is quite high. As long as liquid refrigerant is returning from the ground coil, the orifices have little or no effect. Whenever gas passes through the orifices, the pressure difference between the inlet and return headers is sufficient to force the trapped liquid into the return header. With the full condensing surface then available, all the refrigerant passing through the coil is condensed and only liquid passes through the orifices. In this way the ground coil loops are kept free of trapped liquid, and operate in the most efficient manner. These orifices have been used on several installations.

This method appears to be adaptable only to units using a single reversal valve. The orifices are placed in the return side of the ground coil when it is acting as a condenser. On units having a single reversal valve, the expansion valve connections are made to this side of the coil when it is acting as an evaporator. It is a



(B)

Fig. 8. Record of operation at 29 Foreway Drive, St. Louis County, Mo., installation G

A—During winter 1952-1953
B—During cooling season, summer 1953

very easy matter to bring the distribution tubing from the expansion valve into the ground coil tubes on the opposite side of the orifice from the header, so that the liquid refrigerant bypasses the orifices during heating operation. On units having two reversal valves, the return side of the ground coil when it is acting as a condenser is also the return side when it is acting as an evaporator. Thus, the orifices would be in the suction side of the ground coil during heating operation, which would result in excessive pressure drops. It might be possible with the use of check valves to bypass the orifices during heating operation but this, of course, would be more complicated than the simple installation of the orifices alone.

The ground coils for the four most recent installations are pitched throughout their length so that the refrigerant flows downward as it condenses during cooling operation. These ground coils are used with units having only one reversing valve so that the refrigerant must rise as it evaporates during heating operation. Three of these units have reversal valves on both compressor systems so that the two systems will alternate during cooling operation.

Operating Results

The operation of nine heat pump installations on customer premises was recorded between May 20, 1953, and May 25, 1954. These installations are equipped with demand and kilowatt-hour meters to measure accurately the electric energy consumed and to show the relation of heat pump operation to the company's other load.

Calculations have been made to determine the heat loss of each of these houses during the winter and the heat gains during the summer in accordance with the prescribed practice of the American Society of Heating and Air Conditioning Engineers. The heating season of 1953-1954 was warmer than normal. The summer of 1953 was more severe than usual as indicated by average monthly temperatures.

The electricity consumption during the winter of 1953-1954, as recorded by the metering instruments, was adjusted to that which would have been consumed in a normal heating season for the St. Louis area. The electric energy consumed for summer was not adjusted.

One year's operation of nine heat pump installations cannot be regarded as a conclusive test of operating costs. The homes in which these heat pumps are installed vary in size between 5,300 and 31,930 cubic feet. The calculated winter heat loss, when the outside temperature is zero F and the inside temperature is 70 F, ranges between 40,000 Btu per hour and 157,000 Btu per hour. The houses were occupied by their owners during the entire period. One installation was subject to operating difficulties but in the opinion of the authors, this did not affect the energy consumed.

The habits of the occupants will vary as to temperature desired in the homes during summer and winter. This results in comparatively wide variations of the ratio of kilowatt-hours used. It is known that one installation requires very little summer air conditioning, partly because of its shaded location and partly because of the owner's preference.

The estimates of heat losses based on formulas and procedures of the American Society of Heating and Air Conditioning Engineers are subject to judgment factors as to the use of basement space for living quarters, and infiltration or introduction of outdoor air. It is difficult to determine accurately whether the workmanship of placing insulation was such as to make the insulation as completely effective as assumed. The effect of wind velocities depends on the orientation of the house. The use of lights and electric appliances, cooking and laundry schedules and the frequency of opening and closing of doors and windows can result in operating characteristics which are difficult to reconcile in comparing installations. Table I shows a summary of the obtained. It is planned to obtain additional data and to continue to analyze that now available.

An average house, representative of these nine structures, would contain 15,000 cubic feet and would have a winter heat loss of approximately 85,000 Btu per hour when the outside temperature is zero F and the inside temperature is 70 F.

This average house would use approximately 15,000 kw-hr to operate the heat pump during a normal heating season of 4,584 degree days. The energy required for air conditioning during a summer equal to that of 1953 would require 6,000 kw-hr. The annual cost of year-round air conditioning with the heat pump under St. Louis electric rates would be $367 for 21,000 kw-hr.

Operating characteristics of two installations are plotted, together with weather data and shown in Figs. 7 and 8. Fig. 7(A) and (B) and Fig. 8(A) cover heating operation and Fig. 8(B) covers a cooling season.

Fig. 7 covers the operation of installation *C* during the winter of 1950-1951, which was somewhat colder than normal. During this winter, there was a particularly cold period of 10 days near the end of January. The degree-days for each day, together with the seasonal total up to that day, are shown. The total of kilowatt-hours consumed is accumulated for each day. At the top of Fig. 7(A) the energy consumed per degree-day is plotted. In Fig. 7(B) four daily temperatures are plotted as follows:

1. The ambient ground temperature 4 feet below the surface (determined from the thermocouples at installation A and assumed to be the same throughout the St. Louis area).
2. The saturated refrigerant temperature corresponding to the compressor suction pressure.
3. The saturated refrigerant temperature corresponding to the compressor discharge pressure.
4. The difference between ambient and refrigerated ground temperature.

The latter curve is the temperature difference responsible for the flow of heat from the ground to the refrigerant in the ground coil. It may be noticed from Fig. 7(A) that the curve showing kilowatt-hours per degree-day which is an inverse measure of the efficiency of the operation of the unit, falls off somewhat near the end of the heating season. The reason for this is not known, but it is likely that the owner of this installation, an experienced refrigeration engineer, made some adjustments to the expansion valves or some similar change which resulted in the improved efficiency.

Fig. 8(A) covers the winter of 1952-1953 at installation *G*. This winter was slightly warmer than normal as shown by the inset graph of average monthly temperatures for this winter, the normal winter, and an extremely cold winter. The degree days and kilowatt-hours for each day are plotted, together with the daily ratio of these two quantities. Both degree days and kilowatt-hours are also plotted on an accumulative basis. It is interesting to note the similarity between the degree days and energy curves.

Fig. 8(B) is merely a continuation of (A) showing cooling operation at installation *G* during the summer of 1953. The temperatures this summer were well above normal but somewhat less than those of the two hottest summers on record, 1934 and 1936. This variation is shown on the inset graph. The daily kilowatt-hours, accumulated kilowatt-hours, and daily mean temperatures are plotted on the graph. Although there is naturally a correlation between the daily kilowatt hours and mean temperature curves, the similarity is not nearly as pronounced as that between the kilowatt hours and degree-days' curves on the previous graph.

Some of the heat pumps installed in St. Louis are composed of selected refrigerating equipment installed in the basement of the structure. Others are compactly assembled by local contractors into a cabinet and are of a very neat appearance. There are also a number of heat pumps produced by the factory of manufacturers, although these did not have the benefit of quantity production methods. All of these heat pumps are made up of parts and components used by the industry for air-conditioning and refrigeration purposes. The refrigerant in all cases is *F-12*. It is obvious that none of these installations can be regarded as finished products. When first installed, they have required considerable attention in the way of adjustments and minor changes. However, they are working well.

It must also be obvious that the installed cost of these heat pumps is high and cannot in any way be regarded as a measure or even as an indication of what heat pumps will cost in the future. It has been attempted here to arrive at the cost of these heat pumps, including in this cost all ductwork and air outlets, thermostats and control equipment, the ground coil and all miscellaneous costs, and to arrive at a price of a turn-key job. For what it may be worth, the authors believe this figure to be $850 per horsepower, so that a 10-hp year-round air-conditioning job which would be sufficient for a 85,000 Btu house in the St. Louis area is approximately $8,500. From a commercial standpoint, such a price permits the sale of heat pumps only in special cases. This situation is no different from that of air conditioning in the days before standardized equipment was produced by the manufacturers.

The "prophets" of 1930 were certain that air conditioning would never become widely used in residences because of its high initial cost. It has taken about 20 years to disprove that contention, but during these 20 years, progress was retarded by a severe depression and a world war. It seems certain that the aggressiveness and ingenuity of the manufacturers of the air-conditioning and refrigeration industry will produce a heat pump which will be readily salable. This company's efforts are directed to providing the industry with information, particularly operating characteristics and operating costs, so that the time which so frequently elapses between the inception of a new idea and its wide public acceptance will be reduced to a minimum.

References

1. Climatology as an Aid in Heat Pump Design, **George S. Smith.** *ASHVE Transactions*, American Society of Heating and Ventilating Engineers, Inc., New York, N. Y., vol. 57, 1951, pp. 499–516.
2. Factors Useful in Ground Grid Design for Heat Pumps, **George S. Smith.** *Ibid.*, vol. 56, 1950, pp. 189–208.
3. Intermittent Ground Grids for Heat Pumps, **George S. Smith.** *ASHAE Transactions*, American Society of Heating and Air-Conditioning Engineers, New York, N. Y., vol. 62, 1956, pp. 473–88.
4. Earth Heat Pump Research, Part II, **Estel B. Penrod, R. B. Knight, Merl Baker.** *Bulletin no. 18*, Engineering Experiment Station, University of Kentucky, Lexington, Ky., 1950.
5. University of Kentucky Heat Pump, No. 1, **Estel B. Penrod, Merl Baker, Alfred M. Levy, Paul M. Chung.** *Bulletin no. 30, Ibid.*, 1953.
6. Earth Heat Pump Research, Part IV, **Estel B. Penrod, Paul L. Pfenningwerth.** *Bulletin no. 35, Ibid.*, 1955.
7. The Design of a Heat Pump Buried Coil, **Donald M. Vestal, Jr., Billy J. Fluker.** *Bulletin 139*, Engineering Experiment Station, Agricultural and Mechanical College of Texas, College Station, Texas, 1957.
8. Soil Temperature, Moisture Cailbnt, and Thermal Properties, **Clyde L. Carter.** *Bulletin no. 15*, Engineering Experiment Station, University of Tennessee, Knoxville, Tenn., 1951.

Discussion

George S. Smith (University of Washington, Seattle, Wash.): This paper presents some timely and valuable information on actual experience with heat pumps using ground coils as a heat source. A considerable portion of my research activity in the past 10 years deals with heat sources for heat pumps with the major portion applied to the study of ground coils.[1–3] Many other rather extensive studies have been conducted in other parts of the country.[4–8]

Such studies appear to produce convincing evidence that a properly designed ground grid can be a very satisfactory source of heat where well or spring water are not available

and where cold frosty spells preclude the use of outside air. One must hasten to warn that the area available on the average city lot in many climates will often not supply the heat needed for the house on this lot.

Unfortunately, a number of poorly designed ground coils installed about the time the ground source was proposed, have failed, reducing the interest in this heat source to almost nothing for many years. A few successful reports will greatly aid in revising this interest.

The authors do not present a well-defined analysis of the design of the coils used, though the results appear to be satisfactory, and for this area such information may prove adequate. In general, however, the design of a ground coil should be as carefully made as any other heat exchanger. Thus far few satisfactory methods have been proposed. In references 2 and 3, a rather sketchy method has been prepared which gives fair results. Vestal and Fluker[3] propose a rather straightforward method which checks very well with my results and with successful grids. Needless to say, more work in the field of design is needed.

In general, I believe it is safe to say the depth should be from about 2 feet in very mild climate to 4 or 6 feet in moderate climates, or even 6 to 8 feet in colder climates. The horizontal spacing should be the equivalent of not under 2 feet, and greater if possible, so that each tube will have a minimum effect on the adjacent one. A lesser distance is possible where two tubes are placed on either side of a trench, providing that the trenches are spaced much further. To supply a given amount of heat per season, enough area must be allotted to the ground coil to supply the heat without lowering the temperature to the point where it will not be restored during the following summer. Only enough tubing to remove this heat is needed; more than this would result in excessive first cost.

As fosail fuels become more expensive, we might find a combination of air coil plus a ground coil for the cold spells, when air cannot be readily used. This might solve the problem of a heat source on the average city lot, especially if the ground is used as a heat sump for summer conditioning. Since the travel of heat in soil is rather slow, the ground may prove to be an excellent heat storage medium, provided that the coil is placed 4 to 6 feet deep as a minimum. Of course, such storage precludes the use of earth volume in which shifting underground water is present.

E. A. Freund and G. S. Whitlow: We are grateful for Prof. Smith's interest in our paper, and we certainly agree that the success of ground coil heat pumps depends to a great extent on a carefully designed ground coil, matched to the refrigeration system.

Few utilities are equipped and staffed to do pure research, and our work is intended as a field investigation to determine: 1. Characteristics necessary for a sound heat pump design, 2. Operating costs and practicable coefficient of performance, and 3. Relative merits of different heat sources.

Our investigation indicates that heat extraction rates of 30 to 40 Btu per foot of copper pipe is feasible. Greater extraction rates result in severely depressed suction pressures and resultant inefficient operation. A lesser extraction rate results in inefficient use of tubing and either excessive suction pressure loss or sluggish movement of refrigerant and failure to return oil.

Horizontal spacing greater than 6 feet appears to be unjustified. An experiment of measuring temperatures at various distances from ground coils indicated that there was little noticeable effect beyond 4 feet.

Ground coil depths just below the local frost line are adequate. Heat flow into the ground coil area is insufficient in cold weather to prevent freezing the earth. When the earth around the ground coil is frozen, the unfrozen earth between the coil and the surface constitutes an infinite insulator and no heat can be lost to the surface regardless of its temperature. Because of the infinite mass below and to the sides of the ground coil, the gain in performance by lowering it one, or several, feet would be small, and the cost of so doing would be great.

Moisture migration toward a cold body appears to provide ample moisture at the ground coil when it is used as a heat source. Migration away from the warm coil is detrimental to its operation as a heat sink.

Fig. 4, which is typical of several years' data, shows that the ground coil area temperature recovers completely during the rest periods in the spring and fall. There seems little danger that recovery may be incomplete in other areas, even in more extreme climates.

The earth should be an excellent low-temperature heat storage medium. Simply thawing the frozen earth which develops will provide substantial storage. If shifting or flowing water is available in the ground coil area, storage will not be necessary to obtain high coefficients of performance.

Determining Actual Power Output and Derating Factor of Electric Motors That Drive Oil Well Pumping Units

M. H. HALDERSON
MEMBER AIEE

ELECTRIC MOTORS that drive oil well pumping units cannot deliver an average horsepower output equal to name-plate rating without the effective or rms current exceeding rated current. Expressed another way, motors that drive pumping units must be derated if they are to operate without being thermally overloaded. A method is presented for determining the per cent of rated power output that electric motors can deliver when driving oil well pumping units. This requires a determination of average power output under actual operating conditions. A procedure for determining the average power output is developed that only requires measurements which can be made in the field by engineers and field testers.

Understanding the problem requires a knowledge of the type of load encountered in oil well pumping and what useful information is obtained from the many electric instruments that may be applied to motors driving pumping units. Characteristics of the oil well pumping load and the application of electric instruments are discussed before developing the procedure for obtaining average power output and derating factor. Although the discussion refers to motors that drive oil well pumping units, the procedures presented can be applied to any motor that is subjected to a continuous cyclic load.

Oil Well Pumping Load

The load at the crankshaft of a properly counterbalanced pumping unit is similar to that shown in Fig. 1. Actual curves of crankshaft torque versus time are irregular in various ways, but the load variation and the positions of the load limits with respect to average load are typical. The electric load of the motor does not vary as widely compared to average load as the load at the crankshaft. Inertias of

Paper 58-465, recommended by the AIEE Petroleum Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Petroleum Industry Conference, Dallas, Tex., September 16–17, 1958. Manuscript submitted May 26, 1958; made available for printing July 10, 1958.

M. H. HALDERSON is with the Phillips Petroleum Company, Bartlesville, Okla.

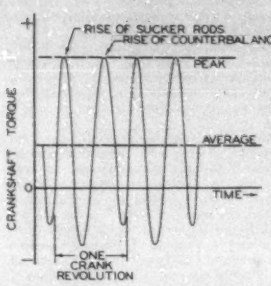


Fig. 1. Load at the crankshaft of a properly counterbalanced pumping unit

rotating parts in the system absorb energy on peaks and give it up on the valleys. The result is that peak motor electric torque referred to crankshaft is less than peak torque at the crankshaft. Peak motor electric torque is referred to crankshaft by multiplying it by the ratio of motor speed to speed of crankshaft.

The load-versus-time relationship is influenced by many factors and is different for each installation. Fig. 2 shows recording wattmeter records for a number of different wells. These are not accurate records of load versus time but they show that there are differences in load characteristics. The ratio of peak load to average load varies over a wide range, being as high as 4 to 1 and as low as 1.8 to 1. The highest peak to average ratios are found where the induction motor driving the pumping unit momentarily becomes an induction generator at the top and bottom of the pumping stroke. During these intervals energy is transferred from the pumping unit through the motor to the power distribution system.

Oscillograph records of power, current, and voltage at the input to electric motors driving pumping units are shown in Fig. 3. The oscillograph records of power show that power is delivered at a frequency of 120 cycles per second. The records of current look like textbook examples of amplitude-modulated radio waves. Note the rise in current that occurs when load goes negative after the peaks.

Reaching general conclusions regarding ratio of peak load to average, motor load carrying capacity, and efficiency, requires that tests be made on many electric motor-driven pumping units. As tests are made on more and more installations, these various items are found to vary over a range of values. The greater the number of tests, the better will be established the limits of the ranges and the average values of the items being checked.

Application of Electric Instruments

A number of electric instruments are available for taking measurements at the input of electric motors that drive pumping units. These instruments include recording wattmeter, indicating wattmeter, recording ammeter, indicating ammeter, watt-hour meter, thermal ammeter, and voltmeter. Special, higher priced instruments that may be used include the oscillograph ammeter, oscillograph wattmeter, and oscillograph voltmeter.

Because of damping and inertia of moving parts, recording wattmeters do not give a true picture of a pumping load. This is also true of indicating wattmeters, recording ammeters, and indicating ammeters. This point is illustrated by Fig. 4. Chart *A* shows power input to a pumping motor as recorded with the pen swinging freely. Chart *B* shows the true peak power on the rise of the rods. This was obtained by pushing the pointer upscale with the side of a screwdriver or pencil held at right angles to the plane of motion of the pointer. The pointer was moved upscale, until the needle no longer kicked off the side of the screwdriver on the rise of the rods, and then was moved back slightly so that the movement of the needle from the side of the screwdriver was barely perceptible by sighting against the white background of the chart. On chart *B*, the power peak on the rise of counterweights is still kicking the needle off the side of the screwdriver about one scale division. Following the peak on the rise of the rods is the valley or minimum that occurs as the crank passes through its bottom position. This was obtained by shifting the zero upscale and pushing the pointer downscale until it barely kicked off the side of the screwdriver as the crank passed through its bottom position. This is shown on chart *C*. The actual peak power input on rise of weights was obtained by shifting the zero to the point shown on chart *D* and again following the procedure of chasing the pointer upscale with the edge of a screwdriver until on the rise of the rods the pointer barely kicked off the sides of the screwdriver. The minimum or valley following rise of weights is shown on chart *E*. This was obtained by shifting the zero upscale and forcing the pointer downscale until it just perceptibly left the side of the

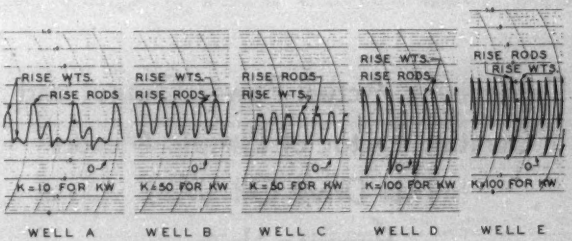


Fig. 2. Recording wattmeter records of power input to motors driving oil well pumping units

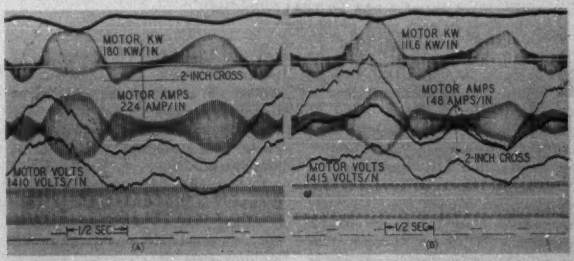


Fig. 3. Oscillograph records of power, current, and voltage at input to motor driving an oil well pumping unit

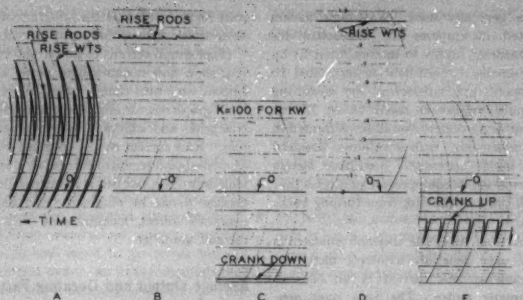


Fig. 4. Illustration of performance of recording wattmeter when measuring input to 5% to 8% slip motor driving pumping unit 31.5 strokes per minute

A—Pen swinging freely
B—Pen pushed upscale to get true peak on rise of rods
C—Pen pushed downscale to get true valley following rod peak
D—Pen pushed upscale to get true peak on rise of counterweights
E—Pen pushed downscale to get true valley following weight peak

screwdriver as the crank passed through its top position. Obtaining actual peaks and valleys by the procedure outlined requires observation of crank position while the well is pumping, in order to identify the given peak or valley. These must be noted on the chart at the time the record is taken.

For the case illustrated in Fig. 4, the horsepower input limits with the pen swinging freely were 72, 0, 94, 36. The peak on the rise of the rods is listed first. By pushing the pointer upscale and downscale to obtain true peaks and valleys, the true input horsepower limits are 112, −65, 126, −19. The over-all power variation with the needle swinging freely was 94 hp (horsepower), whereas the actual power variation was 191 hp. The pumping unit where this test was made was operating 31.5 strokes per minute with a 75-inch stroke, and was being driven by a 60 hp 5% to 8% slip motor. Average power input was 44.5 hp and the ratio of input peak to average was 2.84.

Error in recording limits with the pointer swinging freely is also present at slower pumping speeds, as illustrated by Fig. 5. Horsepower input limits with pointer swinging are 72, 20, 62, 4. By pushing the pointer upscale and downscale actual input horsepower limits are found to be 92, −5, 84, 4. These records were taken with a pumping unit being driven 17.7 strokes per minute by a 5% to 8% slip motor rated 60 hp. Records taken with a 60-hp normal slip motor driving the same pumping unit at about the same speed are shown on Fig. 6. Horsepower input limits as recorded with pointer swinging freely were 92, 10, 84, −2. With pen movement restricted to record true peaks and valleys, the horsepower input limits were 137, −16, 113, −16. The last two examples show that error with pointer swinging is greater when input to normal slip motors is measured.

Recording wattmeter charts taken with the needle swinging may be used to approximate the average power input to a pumping motor. This approximation should not go beyond estimating the average power with the eye. Accuracy of recording is not good enough to justify using a planimeter to get average power.

By pushing the pointer upscale and downscale, recording wattmeters, and also indicating wattmeters, may be used to obtain the actual input power limits to motors that drive pumping units. The power input peaks may be used to determine the state of counterbalance of the pumping unit. A pumping unit is considered properly counterbalanced when the peak input on rise of rods equals peak input on rise of weights.

Indicating ammeters with pointer swinging freely do not give an accurate picture of current versus time when applied to a motor that drives a pumping unit. Actual current peaks are always higher than indicated with pointer swinging. True current peaks can be obtained by pushing the needle upscale as outlined for wattmeters. Recording ammeters do not give any more usable information than the indicating ammeters.

Indicating ammeters may be either the type that are connected in one of the leads to the motor or may be the clamp-on type. The clamp-on ammeter is more generally used because of its ease of application. Some clamp-on ammeters are combined with a voltmeter in the same case.

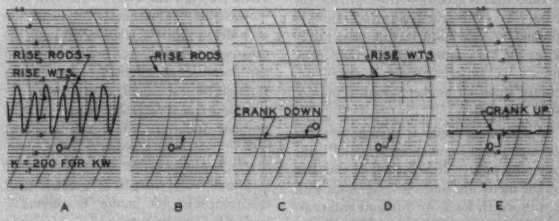


Fig. 5. Illustration of performance as in Fig. 4, driving pumping unit 17.7 strokes per minute

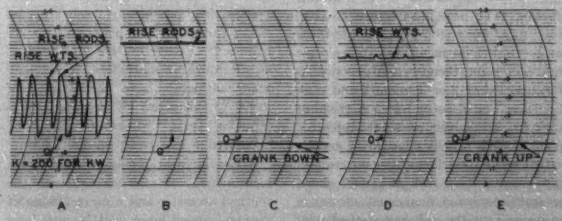


Fig. 6. Illustration of performance as in Fig. 5, when measuring input to a normal slip motor driving same pumping unit as 5% to 8% slip motor for which records are shown in Fig. 5

Clamp volt-ammeter is a term applied to these combination instruments.

When applied to a motor that drives the pumping unit, the pointer of an indicating ammeter moves continuously. If the pumping unit is counterbalanced to some degree, the needle will swing up on the rise of the rods, fall back, and possibly give a small kick upscale as the crank goes through the bottom position then swing up on the rise of the counterweights and swing back to a minimum value, with possibly a short kick upscale as the crank goes through its top position. Observation of the swinging pointer of an indicating ammeter is of value only to determine the state of counterbalance of a pumping unit. The unit is considered properly balanced when the peak current on rise of rods is equal to the peak current on rise of counterweights. Observation of the position of the crank or polished rod is required in order to identify the peak that is being measured. Counterbalance is usually considered satisfactory if the ratio of higher peak to lower peak is less than 1.2.

The indicating ammeter tells an observer nothing about the rms current load of a motor and therefore it cannot be used to determine whether or not an oil well pumping motor is overloaded.

Some clamp-on ammeters have an adjustable pointer stop, a desirable feature built into the meter that permits pushing the pointer upscale by turning a screw or a knob that is accessible from outside the case. To get a true peak, the pointer is pushed upscale until it just perceptibly kicks off the pointer stop when the peak occurs. Counterbalancing is more accurate when an ammeter with a pointer stop is used.

A watt-hour meter is used to measure average power input to motors that drive pumping units. Revolutions of the disk on the meter are timed several times with a stop watch for 2- or 3-minute intervals and a representative average disk revolution per minute is obtained by averaging the results of several 2- to 3-minute counts. The best procedure is to snap the "start" on a stop watch as the mark on the watt-hour meter disk passes some reference point, then to count revolutions for a period of 2 or 3 minutes and snap the "stop" on the stop watch at the time the mark on the disk passes the reference point. Wells usually change to a certain extent when shut down for connection of instruments. The disk should be timed several times before final counts are taken to insure that the well has reached relatively steady operating conditions. Three-phase 480-volt watt-hour meters are generally used. Watt-hour meters rated 15 amperes are satisfactory for measuring inputs to motors rated 20 hp or smaller. Watt-hour meters rated 50 amperes are satisfactory for measuring inputs to motors rated 25 to 75 hp. Watt-hour meters should be calibrated in a meter shop with a rotating standard to insure accuracy. Watt-hour meters should not be equipped with a ratchet that will prevent the disk from turning backwards.

At this time, the thermal ammeter is the only generally available meter for measuring rms current of an oil well pumping motor. The type most generally used is combined with a split-core, multiple-ratio current transformer. These combination thermal-ammeter current-transformer instruments appear to resist damage better than most instruments but they must be calibrated. Error in the readings of new instruments may be as much as 7% or 8% of full scale. A phantom current loader and an ammeter with an accuracy of $\pm 1/2\%$ of full scale are usually used for calibrating thermal ammeters.

One disadvantage of a thermal ammeter is that 50 minutes to one hour is required for the meter to reach correct reading. More accurate readings are obtained if pointer deflection is half-scale or greater. The greater scale deflections can be obtained by providing slack in the motor lead and passing a number of turns through the core of the split-core transformer. The meter constant is then determined by considering both the ratio set on the current transformer and the number of conductors that pass through the core of the transformer.

Measurement of rms current with a thermal ammeter is the only practical way to determine whether or not a motor driving a pumping unit is overloaded. If measured rms current is greater than rated current the motor is thermally loaded beyond its rating.

An indicating voltmeter is used to check the operating voltage, to determine voltage drop on starting the motor, and to determine how well phase voltages are balanced. Voltmeters are also used to locate trouble in controllers. Commonly used voltmeters have an accuracy of $\pm 3\%$ of full scale. Where precision is required, these instruments should be calibrated before being used.

A measurement of voltage along with a measurement of rms current with the thermal ammeter and a measurement of average power input with a watt-hour meter permits calculation of power factor. On 3-phase systems, power factor equals watt input divided by the square root of three times voltage times rms current.

Oscillograph records of power and current show how instantaneous power and current vary with time. Practical use of oscillograph records is limited to providing true peaks and valleys of power and current. Rms current can be obtained from an oscillograph record of current by a very tedious process, but this is a waste of time because it can be obtained much more easily by direct measurement with a thermal ammeter.

Average Output and Derating Factor

Oil producers have procedures for estimating the prime mover output required for sucker-rod pumping installations. Engines are usually applied so that the average power output will be 65% to 70% of the maximum brake horsepower at operating speed. Electric motors are applied to develop a power output less than name-plate rating in order to prevent thermal overload, but there is a question of how much less. When a motor drives a pumping unit, the oil producer needs to know what per cent of rated power output he can expect a given type of motor to develop without being thermally overloaded. Proceeding from this thinking, derating factor is defined as the ratio of average motor output on a pumping load to the output that the motor would develop on a steady load at the rms current measured on the pumping load. When measured rms current is near the full load current rating of the motor, derating factor becomes the factor by which the motor power rating should be multiplied to give the average output that the motor can deliver on a pumping load without being overheated. Obviously, derating factors must be obtained from tests made in the field on electric motor-driven pumping units.

Tests to determine the load-carrying capacity of motors driving pumping units should be taken under conditions where the measured rms current is between 80% to 120% of rated current. A 5% to 8% slip design *D* motor at half-load will have a slip equal to that of many high-torque, normal slip design *C* motors operating at full load. Therefore, design *D* motors operating at light loads will behave more like loaded design *C* motors and derating factors calculated at light loads will be low.

Obtaining derating factor as defined requires a determination of average motor output under actual operating conditions. Average output cannot be measured directly but it can be found by a method

which does not require special instruments or equipment. Briefly, this method is as follows: 1. Connect a watt-hour meter to measure average power input to the motor. 2. Connect a thermal ammeter to measure rms current. 3. From the performance curves of the motor determine the output and efficiency on a steady load at the rms current measured on the test. 4. Divide steady-load output by efficiency to get the input to the motor on a steady load at the measured rms current. 5. Subtract steady-load output from the steady-load input to get motor loss. 6. Subtract motor loss from measured power input to get actual average output. Derating factor is determined by dividing the average output on the test by the output the motor would have developed on a steady load at measured rms current.

The procedure outlined is illustrated by a test where a 60-hp 440-volt 77-ampere 1,170-rpm high-torque normal-slip design *C* motor was driving a pumping unit 18.6 strokes per minute. The performance curves of this motor are shown in Fig. 7. From measurements on test:

Average motor input = 49.4 hp
Rms current = 84.7 amperes

On a steady load at 84.7 amperes: point *A*, Fig. 7:

Output = 66.7 hp; point *B*, Fig. 7
Efficiency = 0.92; point *C*, Fig. 7
Input = 66.7/0.92 = 72.5 hp
Loss = input − output = 72.5 − 66.7 = 5.8 hp

On the test:

Average output = measured input − loss
= 49.4 − 5.8 = 43.6 hp

$$\text{Derating factor} = \frac{\text{Average output on pumping load}}{\text{Output of motor on steady load at rms current measured on pumping load}} = 43.6/66.7 = 0.654$$

The foregoing presents data on an installation where a normal slip motor was operating at a current 10% above rated, whereas average output was only 73% of rated.

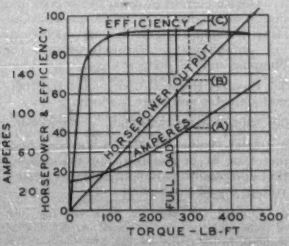


Fig. 7. Performance curves for a 60-hp 440-volt 77-ampere, 1,170 rpm high-torque normal-slip design C motor

One practical use of derating factors is to compare the performance of normal slip motors with that of 5% to 8% slip motors when driving pumping units. A 60-hp 440-volt 1,125-rpm 71-ampere 5% to 8% slip, design *D* motor was installed in place of the 60-hp, normal slip motor for which data are given in the foregoing example. The performance of the 5% to 8% slip motor was as follows when rate of production was within 1% of the same number of barrels per day:

From measurements taken on test:

Average motor input = 49.4 hp
Rms current = 60.0 amperes

On a steady load at 60.0 amperes:

Output = 50.1 hp
Efficiency = 0.896
Input = 55.9 hp
Loss = 59.9 − 50.1 = 5.8 hp

On the test:

Average output = 49.4 − 5.8 = 43.6
Derating factor = 43.6/50.1 = 0.87

The foregoing is a typical example of cases where a 5% to 8% slip motor is considerably underloaded when replacing an overloaded normal slip motor of the same rating. The results are in agreement with data previously presented which showed that 5% to 8% slip motors can develop 30% to 35% more average power on a pumping load without being overloaded.[1] In the example cited, a 50-hp 5% to 8% slip motor would have been satisfactory where the 60-hp normal-slip motor was overloaded.

The load variations were less severe than normal on the installation used in the example, which accounts for the derating factors being higher than average. Tests reported in the reference showed average derating factors to be 0.58 for normal slip motors and 0.80 for 5% to 8% slip motors, where rms currents ranged from 80% to 121% of rated.

Voltage was 450 at the terminals of the two 60-hp motors for which test data are given. This, along with the measured power input and measured rms current, permitted calculation of power factor. Power factor was 0.79 for the 5% to 8% slip motor and 0.56 for the normal slip motor. The higher power factors found with 5% to 8% slip motors permit savings in capacitors, distribution lines, and transformers.

Accuracy

The procedure for obtaining average output is accurate if voltage on the test is equal to rated voltage of the motor. Error in average output determined by the method outlined will be less than 1% if terminal voltage is within 7½% of rated, and if calculated output is greater than 45% of rated. For these conditions, the accuracy of the method is as good as that of the measurements and the performance curves that manufacturers usually furnish.

Error in calculated output will increase as motors are more lightly loaded, and the more operating voltage deviates from rated. Error can be held to 1% or less if core loss is obtained in addition to the information usually furnished with performance curves. Motor loss includes I^2R (current² resistance) loss, friction loss, and core loss. Friction is a small part of the total loss and may be considered constant. I^2R loss is determined by rms current which is measured. Core loss is made up of two parts, one due to hysteresis which varies as the 1.6 power of voltage and one due to eddy currents which varies as the square of voltage. Hysteresis loss will usually be between three to four times eddy-current loss. A close approximation for adjusting core loss for changing voltage is to assume that core loss varies as the 1.75 power of voltage. Therefore, if voltage is 10% higher than rated, core loss will be increased to 1.1 to the 1.75 power times core loss at rated voltage.

Whether or not adjustment of core loss is necessary to have desired accuracy in calculated output depends on the motor design as well as per cent load and deviation of operating voltage from rated. Examination of characteristics of seven motors showed that there are differences in the need for adjusting core loss. Ranges found for minimum values of average

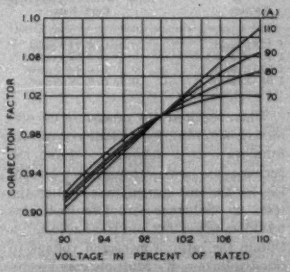


Fig. 8. Factor for correcting motor power output at a given current and rated voltage for deviation in voltage from rated. A = power output at given current and rated voltage in per cent of rated

Well _____ Date _____ By _____

MOTOR

HP _____ Volts _____ Phase ____ Cycles _____ Amps _____ °C Rise ____

RPM _____ Make _____ Style, Type, or Model _____

Design _____ Serial Number _____ Frame _____

PUMPING UNIT

Make and Model _____ C-Wt. Setting _____

Measured Stroke _____ C-Wt. Number _____

DRIVE

Pumping Unit Sheave P.D. _____ Motor Sheave P.D. _____

No. V-Belts _____ V-Belt Section _____

TEST DATA

RUNNING VOLTAGE: Ph. 1 _____ Ph. 2 _____ Ph. 3 _____

CURRENT PEAKS: Ph. 1 R ___ W ___ Ph. 2 R ___ W ___ Ph. 3 R ___ W ___

RMS CURRENT (K =) Ph. 1 _____ Ph. 2 _____ Ph. 3 _____

Strokes Per Minute			
Time of Day	No. of Strokes	Min.	SPM

Motor Input (K =)			
Time of Day	Disc Rev.	Min.	Disc RPM

Fluid Level	
Shot No.	Time of Day
1	
2	
3	
4	
5	

Dynamometer Cards	
K =	Lbs/Inch
Card No.	Time of Day
1	
2	
3	
4	
5	

Producing Rate	
Bbls/Inch =	
Time of Day	Gage

Fig. 9. Example of data sheet for field tests on electric motor-driven pumping units

power output to hold error in calculated output below 1% without correcting core loss are given by the following table.

Voltage	Ranges of Minimum Power Outputs for Less Than 1% Error Without Correcting Core Loss
1.125×rated	0.39 to 0.74×rated
1.10 ×rated	0.31 to 0.59×rated
1.075×rated	0.23 to 0.44×rated
1.05 ×rated	0.15 to 0.29×rated

When required, the procedure for correcting motor loss for deviation of voltage from rated is as follows: 1. From the performance curves determine total motor loss for the rms current measured on the test. 2. Subtract core loss from this total loss to get I^2R and friction loss. 3. Obtain the ratio of measured voltage to rated voltage. 4. Multiply core loss by this ratio raised to the 1.75 power to get core loss at measured voltage. 5. Add adjusted core loss to I^2R and friction loss to get total loss at the voltage measured on the test.

Calculating derating factors requires a determination of the output that the motor would deliver on a steady load at the rms current measured on the test. Motor output at a given current varies to some extent with voltage. This should be taken into consideration if derating factors are to be obtained with the greatest possible accuracy.

The degree of change in motor output at given current with change in voltage depends on the per-cent load. This is illustrated by Fig. 8 which gives correction factors by which power output at rated voltage and a given current should be multiplied to give output at voltages ranging from 90% to 110% of rated. The curves shown are averages of several motors of different types and sizes. Use of the chart in Fig. 8 is as follows:

1. From the performance curve of the motor, determine the power output at the current equal to rms current measured on the test.

2. Divide power output determined in step 1 by rated output and multiply by 100 to get output in per cent of rated.

3. Divide terminal voltage measured on the test by rated voltage and multiply by 100 to get operating voltage in per cent of rated.

4. Enter the chart at the bottom at the correct per-cent voltage and move upward to the correct per-cent load by interpolating between the curves for the various loads.

5. Move horizontally to the left to obtain correction factor.

6. Multiply output determined in step 1 by the correction factor to find the output at measured voltage.

The characteristics of some particular motors may differ enough from those used to obtain the curves on Fig. 8 to justify preparing a set of similar curves that apply specifically to these motors. Users should be able to obtain information on this from the manufacturer of the motors that they are using.

Field Testing

A thermal ammeter and watt-hour meter will supply the essential information for calculating motor output and derating factor. A voltmeter will provide measurements needed to calculate power factor, to make corrections for deviation of operating voltage from rated, and to determine balance of phase voltages. Current peaks and the state of counterbalance of the pumping unit may be obtained with a clamp-on ammeter. Improper counterbalance lowers the derating factor and a pumping unit should be reasonably well counterbalanced if the results are to be representative of good operating practice.

An aid in field testing is a data sheet which provides blanks for recording data to be taken on a test. An example of a data sheet for field tests on electric motor-driven pumping units is shown on Fig. 9.

The test procedure in the field is as follows:

1. Obtain current peak on rise of rods and on rise of weights in one phase to determine state of counterbalance for normal operating conditions. Recounterbalance the pumping unit if necessary.

2. Connect the watt-hour meter and thermal ammeter and start the unit.

3. Record name-plate data of the motor.

4. After timing the disk of the watt-hour meter at 4- or 5-minute intervals shows that well conditions have stabilized, measure voltage on all three phases.

5. Obtain the current peak on the rise of rods and the current peak on the rise of weights in all three phases.

6. Time the disk for two or three $2^1/_2$- to 3-minute intervals to obtain an accurate average disk rpm. Average motor power

input is calculated from the following equation:

$$\text{Average hp input} = \frac{\text{disk rpm} \times \text{disk } K \times 60 \times 1.34}{1{,}000}$$

7. Read the thermal ammeter, being sure that the instrument has been carrying motor current for at least 50 minutes. Record the current transformer ratio and the number of conductors passing through the core of the split-core transformer, if used, so that the proper scale factor can be determined.

8. If the current peaks on rise of rods and rise of weights are found to be 5% or more different respectively in the different phases, it is a good idea to check rms current in a second phase. Rms current in the third phase can then be estimated assuming that rms current varies in the same proportion as the average of the current peaks for the various phases. Over-all average rms current is then found by averaging the rms currents for the three phases.

The need for care in taking measurements cannot be emphasized too greatly. Final measurements should not be taken while well conditions are changing. The disk of the watt-hour meter should be timed at 4- to 5-minute intervals prior to taking final measurements, to be sure that conditions are stabilized. This is necessary to insure that the different measurements will be for the same pumping conditions. When in doubt, repeat measurements, particularly the timing of the disk of the watt-hour meter.

Efficiencies

Much more information is needed on efficiencies of pumping units under actual operating conditions. When driven by an electric motor, the efficiency of a pumping unit is determined as follows:

1. Connect a watt-hour meter and thermal ammeter to measure average input and rms current.

2. Install a polished rod dynamometer to measure well load.

3. Follow the procedure described to insure stable well conditions, recognizing that at 50 minutes to one hour is required for the thermal ammeter to give rms current.

4. Measure voltage.

5. Time the disk of the watt-hour meter for 2½ to 3 minutes, during which time let the stylus of the dynamometer trace the dynamometer card continually for at least one minute.

6. Read the thermal ammeter.

7. Calculate average power output of the motor by the method outlined.

8. Calculate polished rod horsepower from the dynamometer card.

9. Calculate the efficiency of the pumping unit by dividing polished rod horsepower by average motor output.

Pumping unit efficiencies can be plotted against polished rod horsepower in per cent of nominal horsepower rating of the speed reducer to establish an average curve of efficiency versus per-cent load.

The operating efficiency of motors that drive pumping units can be determined by obtaining average power output and dividing by measured power input. Efficiency determined in this way should not be used to compare normal slip with 5% to 8% slip motors because it does not include the effect of motor slip on the behavior of the entire pumping system. Energy input per barrel of liquid lifted is needed to judge the efficiency of various types of motors when applied to drive-pumping units. A 5% to 8% slip motor may have a point lower efficiency than a normal slip motor when driving a given pumping unit and, even so, may use less kilowatt-hours per barrel of liquid lifted. Comparisons of system efficiency require accurate tests of motor power input and corresponding producing rate with the two types of motors when pumping a given well. Such tests must be made on wells where pumping conditions will stabilize. Tests should be made with the same pump size and stroke and at nearly the same strokes per minute. Fluid level should be checked by some means such as a sonic instrument to assure that the net lifts are close to the same. A noticeable change in net lift, when a well is pumped at a given stroke and strokes per minute, will indicate a change in well conditions or a change in condition of subsurface equipment.

Program for Progress

Procedures have been outlined for determining average output, derating factor, and power factor of motors that drive pumping units; for obtaining the efficiency of pumping units; and for comparing over-all efficiencies of sucker-rod pumping systems when different types of motors are applied to drive the pumping units. Reliable information on these items is scarce and, without it, costs to equip pumping units with electric motors are higher than they need be.

The first step to obtain needed information is to provide field production offices with a thermal ammeter, a clamp-on ammeter, a voltmeter, and watt-hour meter. These instruments will provide all essential electric measurements. A polished rod dynamometer must be added if test work is to include determination of efficiency of pumping units. The next step is to train personnel how to use the instruments safely and how to obtain accurate and reliable data. The third stop is to have the instruments used to obtain the information needed. If such a program were to be followed aggressively, the right answers would become generally known in a year or two and the job of equipping pumping units with electric motors could be accomplished at the lowest cost.

Reference

1. Selection of A-C Motors for Driving Oil Well Pumping Units. M. H. Halderson. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2078–82.

Discussion

J. K. Howell (Westinghouse Electric Corporation, St. Louis, Mo.): The author's years of experience covering the selection and use of instruments, methods of testing, and obtainable information should be a tremendous help in establishing more clear-cut application procedures. It would be of interest to check the "Push-the-Pointer" method of obtaining peak currents, etc., with an electronic oscillograph.

The method of determining derating factors, outlined in the paper, utilizes a minimum of instruments, and those that are desired for other power tests of a pumping motor; however, certified performance curves are required. By the addition of a phase-shifting transformer, the watt-hour meter may be used to determine the vars the motor is drawing. Knowing the watts, vars, and voltage, the average current can be calculated. The derating factor is then obtained by dividing the average current by the rms current without the use of performance curves.

The NEMA (National Electrical Manufacturers Association) designs *B* and *C* general-purpose motors have a service factor of 1.15. NEMA design *D* general-purpose motors have no published service factor; however, any particular design *D* motor may have some overload capability at rated conditions, depending on the particular design. It is felt that this should be remembered in evaluating high-slip and normal-slip motors. However, the margin in derating factors of the two designs appears to justify the high-slip motor even if the service factor and price differential of the two designs are taken into consideration.

Another point regarding pumping unit efficiency and motor derating should be stressed: the effect of speed of pumping. As speeds are increased, a point of no return is reached.

Mr. Halderson suggests a program of progress in the application of electric motors to walking beam pumping units. It certainly would save considerable engineering in applying motors to this load if universally accepted derating factors could be worked up. Since the derating factor is a function of the pumping speed, the information should be in the form of curves of derating factor versus inches per minute of the polished rod travel for the pumping units of various sizes.

In the application of electric motors to an oil well, we start with the well and pumping unit, and apply an electric motor. Thus, it seems to be more proper to apply to the average calculated load a cyclic load factor to obtain the proper motor size, than to start with a given motor size and derate it to determine the average load it will carry. This is a trivial point, the main job being to establish factors that would be universally accepted.

C. H. Eckel (Phillips Petroleum Company, Bartlesville, Okla.): First, I want to congratulate the author on an excellent paper. He has described a much improved method for testing oil well pumping motors. Furthermore, we are indebted to Mr. Halderson for the work he has done during the past 10 years in studying and reporting on the performance of oil well pumping motors.

The section entitled "Application of Electric Instruments" is very important, as it specifies and instructs regarding the proper instruments and their use in testing design *C* and design *D* motors. Of special importance in this section is the report on his discovery that the damping and inertia of the moving parts of a recording wattmeter prevent the instrument from recording the maximum and minimum peaks that actually exist on a pumping cycle. He has explained a simple method whereby the recording needle can be manipulated so that these actual peaks will be recorded. It is important that the peak kilowatt inputs per cycle be known so that accurate values of performance can be obtained. As the variation of the motor from synchronous speed under load conditions is of prime importance in comparing *C* and *D* motors, it is recommended that a recording tachometer to measure instantaneous motor rpm be added to the list of the instruments that is given.

The design *C* motor has been bought in larger quantities for oil well pumping because it costs less and is 2% to 3% more efficient than a design *D* motor at constant load. It has been thought generally that cyclic loading would not alter this efficiency advantage. The few tests that have been made on design *D* motors indicate this is not so. It is found more pronounced for the higher pumping speeds and greater depths.

In the section "Average Output and Derating Factor," the author compares the performance of a 60-hp design *C* motor with 2.5% slip (1,170 rpm full load) with a 60-hp design *D* motor with 6.2% slip (1,125 rpm full load). The two motors were interchanged on the same well and, consequently, operated the same cyclic load. Data were taken with the instruments previously listed and explained. Results measured with a watt-hour meter indicated that both motors had the same kilowatt input, while measurements made with the thermal ammeter indicated a much higher kilovolt-ampere input to the *C* motor. Because the watt-hour meter has a rotating disk which has some inertia in it and because the cyclic load peaks on the design *C* motor were much higher than those on a design *D*, it is suggested that in similar tests the watt-hour meter be checked for accuracy under cyclic load conditions. This could probably be best accomplished by testing a number of motors simultaneously and having a watt-hour meter on each individual motor, and one on the line supplying the branches to the individual motors. If the individual wattmeters are reading accurately, their sum should equal that of the master watt-hour meter. It is obvious that the *C* motor with 2½% slip is not as satisfactory for the oil well pumping installation where these tests were made.

Now, here are two questions that I feel need to be answered:

1. Are the two 60-hp motors in this test typical of the motors that manufacturers are furnishing for design *C* or *D* motors?
2. Would a design *D* motor of, say 7% or 8% slip show lower cyclic peaks and have a better derating factor?

If the answer to both of these questions is yes, then extensive tests should be made in the manner that Mr. Halderson has proposed in his paper until it is definitely known what slip and other characteristics should be the design criteria for an oil well pumping motor.

When these optimum characteristics are established, they should provide a firm technical base for an oil well pumping motor standard by AIEE or American Petroleum Institute and eventually by NEMA.

M. H. Halderson: The comparison suggested by Mr. Howell of peak current or peak power as obtained by an oscillograph with the peak obtained by pushing the pointer on an ordinary recording or indicating instrument would be not only interesting but desirable.

The statement is made that certified performance curves are required for determining derating factor by the method outlined. Certified performance curves, no doubt, are required to obtain the most accurate results. What errors might result from applying typical curves to off-the-shelf motors is a question that should be answered by the motor manufacturers.

A procedure for determining derating factor is suggested which involves determining average current from a measurement of vars, watts, and volts. Average current is then divided by rms current to get derating factor. An important consideration is that users want a derating factor that tells how much power output an electric motor can average on an oil well pumping application without being thermally overloaded. The ratio of average current to rms current is deficient because it is not related to the ability of the motor to deliver power. Work where average current was determined from measurements of watts, volts, and power factor showed that ratio of average current to rms current gives derating factors that are too high. For example, the derating factor was 0.654 for the normal slip motor used as an example in the paper. The ratio of average current to rms current on this test was 0.87, which is obviously too high. Part of the explanation for the ratio of average current to rms current being too high is that current rises when power goes negative.

A technical question concerns just what current is calculated from a measurement of vars, watts, and volts; or from a measurement of watts, volts, and power factor. It is not obvious that currents so calculated are true average currents.

Even though NEMA does not list a service factor for design *D* motors, a 40 C (degrees centigrade) rise design *D* motor has overload capacity. The maximum allowable temperature for class *A* insulation will not be exceeded until the temperature rise by thermometer exceeds 50 C. An approximation is that temperature rise increases as the square of the load. On this basis a 40 C rise design *D* motor would have a service factor of 1.12. Motor manufacturers should answer the question as to what service factor design *D* motors have irrespective of NEMA's failure to recognize that 40 C rise design *D* motors have overload capacity.

How derating factor varies with pumping speed and inch-per-minute polished rod travel is a question on which there are differences of opinion, and there is practically no information based on the type of tests that would have to be made to settle the question. The determination of application factors, for both normal-slip motors and for 5% to 8% slip motors over wide ranges of pumping speed and inch-per-minute polished rod travel, are needed to get the answer.

A suggestion is to use cyclic load factor instead of derating factor. One factor is the reciprocal of the other, which allows the user to use whichever he prefers.

Mr. Eckel suggests adding a recording tachometer to the instruments listed for testing motors that drive oil well pumping units so that the instantaneous motor speed can be obtained. There may be occasions where knowledge of the speed limits will be of some use. This might be in studies to determine the effect of changing the amount of flywheel effect in the system. However, there is doubt that ordinary recording tachometers would accurately record peak and minimum speeds of motors that drive pumping units. Probably the best way to get actual maximum and minimum speeds is to obtain the peak and minimum power input with a wattmeter by pushing the pointer up and down scale and then reading speed from the motor performance curve.

Doubt is raised regarding the accuracy of a watt-hour meter for measuring power input to motors that drive oil well pumping units. This is a question which may be difficult to settle. A method is suggested which would require accurately determining line losses. Another possible test would be to connect two watt-hour meters of different types, simultaneously to measure power input to given different types of motors. One meter could be a 2-element 2-disk type and the other a 2-element single-disk type. Prior to such a test both meters would have to be calibrated with a rotating standard. Agreement of the two types of meters would in-

dicate their ability to measure power accurately on a continuously varying load. Disagreement would show need for further investigation. One conclusion has been that watt-hour meters can measure accurately under varying load conditions. For example, the statement is made: "A commercial meter, however, is equipped with magnetic damping, which faithfully gives a damping torque proportional to speed under any conditions. It can be proved that with this 'perfect' damping, the failure of the disk to come up to speed initially is exactly made up for by coasting when the load is cut off. There is no meter error."[1]

In reply to Mr. Eckel's two questions:

1. The two 60-hp motors used for the examples are typical design *C* and design *D* motors.

2. The design *D* motor used in the example had a full load slip of 8½%. A motor with a slip of 7% or 8% would be expected to have somewhat lower peaks and a slightly higher derating factor.

REFERENCE

1. A WATT-HOUR METER DEMONSTRATOR, A. D. Moore. *Electrical Engineering*, Oct. 1958, "Watt-Hour Meter Plays Fair," p. 936.

Field Testing of D-C Air Circuit Breakers on a High-Capacity System

P. L. HARTSOCK, ASSOCIATE MEMBER AIEE
T. J. SCULLY, ASSOCIATE MEMBER AIEE
V. N. STEWART, MEMBER AIEE

AN unusual opportunity for testing d-c air circuit breakers on a high-capacity system has been made possible with use of the mercury-arc rectifier system at Alcoa's Rockdale Works. The system had a preliminary short-circuit current calculation of 880,000 amperes but the actual available current was determined at the test point, located mid-way on the d-c bus, to be slightly over 400,000 amperes. The maximum available current was obtained by connecting in parallel four potline busses without a short-circuit current-limiting test resistor. Under this extreme condition, a peak current of 220,000 amperes flowed through the breakers. These current values are greater than the interrupting rating listed in the standards for semi-high-speed d-c breakers normally used for high-capacity d-c application. However, sufficient current-limiting action of the interrupter was obtained and the clearing time was short enough to maintain a continuous power flow to the aluminum potline. Being the first known tests presented since the new rating structure was adopted, the results indicate that the standard[1] is a conservative rating when used as the guide.

Paper 58-225, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted November 6, 1957; made available for printing June 5, 1958.

P. L. HARTSOCK is with the Aluminum Company of America, Warrick, Ind., T. J. SCULLY and V. N. STEWART are with the General Electric Company, Philadelphia, Pa.

The authors gratefully acknowledge the assistance of A. Schmidt, Jr., and W. G. Abbott of the General Electric Company, and of the engineers of the Aluminum Company of America (Alcoa), Rockdale Works, in carrying out the successful test program.

System and Test Arrangement

The system which was combined into a single operating unit to supply the high capacities of the test program is composed of transmission feeders and power conversion units which normally function as separate sources for each of the four potlines. The potline sections are composed of 12 mercury-arc rectifiers having a name-plate rating of 6,000 amperes and 750 volts. This gives 72,000 amperes output in each section and a total name-plate continuous current of 288,000 amperes for the combined system. With such a large capacity, it is reasonable to consider this combined operation as an unusual testing opportunity. The entire system is represented by the 1-line diagram of Fig. 1.

The power conversion units are not shown in the test circuit diagram of Fig. 2, but the blocks of available d-c power are equal in each section of d-c bus between the tie breaker. Several points on the system besides those needed for the breaker test were measured with an oscillograph as part of an over-all system study. These measurements served to establish proper constants and were a basis for improving the technique in system calculations. These calculations are developed in two other papers.[2,3]

The location of the d-c bus in an areaway between the power conversion and processing buildings provides an ideal setup for the testing. The space was situated in such a location as to avoid interference with operations and with sufficient area to handle all necessary equipment and bracing materials. It should be well recognized that the forces in the closely associated conductors at these current levels were extremely high and necessitated considerable bracing before the entire system could be short-circuited. Part of the setup arrangement is shown in Fig. 3.

The magnitude of the available short-circuit current was controlled by two methods; first, series resistance in the test breaker circuit, and second, parallel combinations of d-c busses. Test data on the various combinations and the available currents are given in Table I.

D-C Air Circuit Breaker

The test breaker, shown in Fig. 4, is a single-pole heavy-duty stationary-mounted air circuit breaker having a rating of 8,000 amperes. Silver main contacts for current carrying and intermediate and arcing contacts for the interrupting fuction are used in conjunction with an open-type interrupter. Further, the breaker is used as either a completely open mounting in switchboards or as metal-enclosed equipment having an open top through which the interrupter gases are vented. Closing of the breaker is accomplished by the d-c solenoid

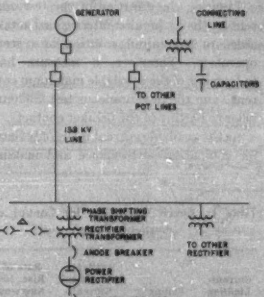


Fig. 1. One-line diagram of Alcoa's mercury-arc rectifier system

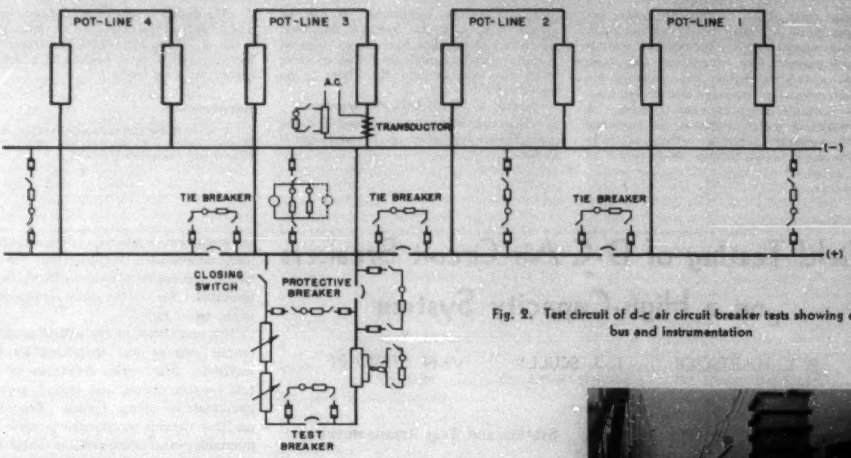


Fig. 2. Test circuit of d-c air circuit breaker tests showing d-c bus and instrumentation

Fig. 3. Test setup at Alcoa's Rockdale Works

Fig. 4. Type MC-5 d-c air circuit breaker during test program

operating through a trip-free mechanism.

Tripping was accomplished by a shunt trip where equivalent high-speed operation was desired or by normal direct-acting overcurrent trip for semi-high-speed breaker performance. Breaker tripping by both methods was used during the tests and is indicated in Table II. In the test with maximum system available current, the direct-acting overcurrent trip was used to demonstrate performance of the general application.

Determination of Available Short-Circuit Currents

Limiting the duration of the test current by the d-c breaker interrupting action aided in maintaining continuous system operation. Therefore, a measure had to be made of the available maximum current from the instantaneous test currents prior to parting of the breaker contacts.

As the available current is the only basis for comparing performance and making applications, this value along with the rate of rise was determined for the range of short-circuit levels tested.

A mathematical extrapolation of the test currents leads to a measure of the available current of the test circuit. The d-c transient is assumed as an *R-L* circuit with a steady voltage applied.

The current rise in an *R-L* circuit of this type is:

$$i = I(1-\epsilon^{-t/T}) \tag{1}$$

where

t = time from closing circuit
T = time constant
I = maximum available current
i = instantaneous value of current

The values of I and T were obtained by taking two readings of the test current (i_1, i_2) at time intervals which are in a 2-to-1 ratio (t_1, t_2).

$$\frac{i_2}{i_1} = \frac{1-\epsilon^{-2t_1/T}}{1-\epsilon^{-t_1/T}} \tag{2}$$

Let

$$a = \text{exponential}\left(\frac{-t_1}{T}\right) \tag{3}$$

Then

$$\frac{i_2}{i_1} = 1+a \tag{4}$$

and

$$a = \frac{i_2-i_1}{i_1} \tag{5}$$

$$I = \frac{i_1}{1-a} \tag{6}$$

Table I. Circuit Conditions for Control of Available Power

Current-Limiting Resistance	Potline No.	Available Short-Circuit Current, Amperes	Rate of Rise, 10^6 Amperes per Second
F	3	85,000	9
E	3	140,000	10
C	2 and 3	225,000	10
B	3	210,000	9
A	3	230,000	12
None	1 through 4	408,000	13

Table II. Field Tests on Heavy-Duty D-C Breakers

Test No.	Trip Means	Available Short-Circuit Current, Amperes	Peak Current, Amperes	Total Clearing Time, Milliseconds	Peak Arc Voltage
162	shunt	85,000	49,400	30.7	860
163	shunt	140,000	81,700	26.8	1,700
164	shunt	140,000	81,700	29.8	1,340
165	direct-acting	140,000	86,500	58.0	
166	direct-acting	140,000	84,500	54.2	750
167	direct-acting	210,000	142,000	61.5	1,380
168	direct-acting	210,000	151,000	50.8	1,700
169	direct-acting	210,000	141,000	49.2	1,700
170	direct-acting	225,000	160,000	65.5	1,550
171	direct-acting	210,000	149,000	50.5	1,670
175	shunt	210,000	121,000	26.8	1,340
176	shunt	210,000	126,000	29.5	1,420
177	shunt	210,000	126,000	30.0	1,380
178	shunt	210,000	123,000	28.3	1,380
183	shunt	230,000	151,000	34.3	1,240
193	direct-acting	*	186,000	54.0	1,270
194	direct-acting	408,000	220,000	55.8	1,360

* Resistor C at half value with approximately 280,000 amperes available.

$$T = \frac{-L}{\ln a} \qquad (7)$$

Equations 6 and 7 are used to calculate values which are compared with the test current measurement before the breaker opens and modifies the system conditions. By an iteration process, the calculated values of I and T can be brought into close agreement with the observations and will give the ultimate maximum available. The maximum available current (Fig. 7) was calculated to be 408,000 amperes with a time constant of 32.9 milliseconds.

Breaker Field Test Performance

In field testing on the multisource system of Alcoa's Rockdale Works, the planning had to be directed toward continuous operation of the potlines while exploring the maximum available short-circuit currents. Therefore, an approach to give increasing currents and to use control of the tripping time was adopted with relatively good success and limited the loss of the potline to a few occasions.

Peak currents through the breaker ranged from 49,000 to 220,000 amperes at 650 volts during the test. The maximum total clearing time using a direct-acting overcurrent trip varied from 0.049 to 0.065 second. These tests with information on the available short-circuit currents and initial rates of rise are shown in Table II. Oscillograms which are typical of the performance of the breaker are shown in Figs. 5 through 7.

As the maximum requirements of the system are beyond the existing standard with respect to interrupting rating, a direct comparison cannot be made. However, it is discussed in the paper under "Breaker Standards."

System Measurements

One of the basic advantages of this extensive test program was the evaluation of the true available short-circuit currents and the response of the system to the imposed faults. Further, it established the actual constants that are associated with the high-capacity d-c systems. In this respect, more points of instrumentation were included in the testing than would be necessary to carry out the field test of a d-c breaker. These points are shown in the test circuit diagram of Fig. 2 and measurements were made to record the changes in the parallel paths including both the d-c circuits and 13.8-kv a-c lines.

Table III. Regulation of Various Parts of System During Maximum Short-Circuit Tests

Location	Measured	Change Between Load and Short Circuit Based on Load, Per Cent
Potline bus	1 voltage, d-c	9.5
	2 voltage, d-c	24.0
	3 voltage, d-c	29.0
	4 voltage, d-c	27.0
13.8-kv potline	1 voltage, a-c	19.0
	1 current, a-c	19.0
	2 current, a-c	47.0
	3 current, a-c	157.0
	4 current, a-c	95.0

The voltages and currents on the 13.8-kv lines were recorded as outputs of a full-wave 3-phase rectifier. The use of the 3-phase rectifier made it possible to record all three phases with one oscillogram trace. The percentage change of the voltages and currents in various parts of the system for the maximum fault condition is contained in Table III.

The division of short-circuit currents based on the normal 13.8-kv-line load currents shows a range of contribution from 19% to 157%. Analysis of this contribution ratio checks very close the 220,000 amperes measured through the breaker. This division is understandable from an examination of the d-c fault location. Although potlines 2 and 4 were on either side of the d-c bus section having the test short circuit, the difference in the bus resistance and inductance to the point of test caused potline 4 to contribute twice

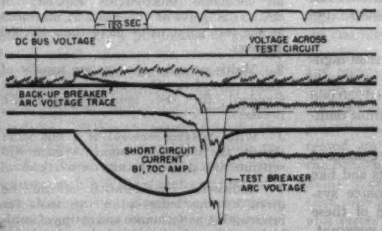


Fig. 5. Oscillogram of short-circuit test with fast tripping

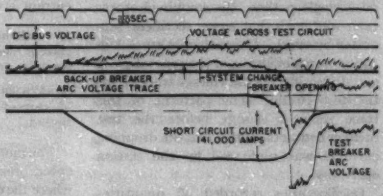


Fig. 6 (right). Oscillogram of short-circuit test showing change before breaker opening

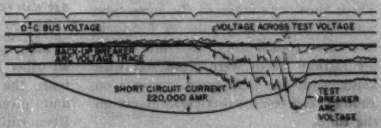

D-C BUS VOLTAGE
VOLTAGE ACROSS TEST VOLTAGE
BACK-UP BREAKER ARC VOLTAGE TRACE
SHORT CIRCUIT CURRENT 220,000 AMP.
TEST BREAKER ARC VOLTAGE

Fig. 7 (right). Oscillogram of short-circuit test with maximum available current from entire mercury-arc rectifier system

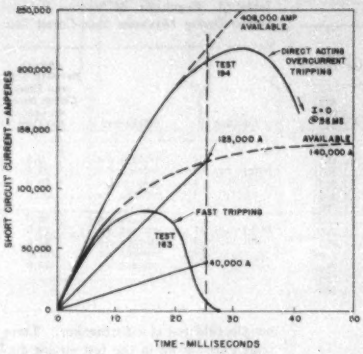


Fig. 8. Comparison of breaker performance with breaker standard

as much current as potline 2. The line furthest from the test location only contributed 11% of what was received through the nearest line.

Connection of the test breaker across the d-c bus put the breaker in parallel with the cell line. Although the aluminum potline has a back electromotive force in a direction to oppose normal current direction, the contribution of the cell to the short circuit was small due to the high inductance. Measurement of the cell line current indicates that the load current was reduced approximately 22% during the maximum fault current test.

High-capacity d-c systems come into existence by combining several sources in parallel. Each source has air circuit breakers with overcurrent tripping devices. The setting of the trip device provides a limit for each branch current. Since the d-c testing was performed on a working system, the trip settings of the cathode breakers of the rectifiers were increased during the program to give maximum trip setting. In high-current tests on a single potline, the combined operation was such that some rectifiers separated by overload from the system. Fig. 6 gives some evidence of the separation of certain rectifiers in the voltage trace across the test circuit. The patterns of this trace show a change before the test breaker contacts parted, and the dropping of units was confirmed by the station operators.

Oscillographs recorded 25 quantities during the high-current d-c testing. Three films were taken at the same time with the same timing pulse at 1/2-cycle interval exposed on each film. The oscillogram traces were organized with voltages on one oscillogram, currents on another, and test breaker performance on the third.

The breaker performance oscillogram, Fig. 7, recorded the short-circuit current, d-c bus voltage of potline 3, voltage across the test circuit, and voltages across the test and protective breakers.

Breaker Standards

Present-day standards and the application guides[1,4] relate the interrupting performance of high-speed and semi-high-speed d-c air circuit breakers to the breaker tripping speed and the rate of rise of the available system current. The experience and tests with large short-circuit duty such as reported in this paper would normally be considered as a semi-high-speed application. However, the rate of rise compared with the standard indicates limits requiring high-speed breakers unless the rules with permissive damage are followed. These rules state that a semi-high-speed d-c air circuit breaker with instantaneous trip devices may be applied in circuits in which the magnitude of maximum available fault current is either:

1. Unlimited for initial rates of rise of current below 5 million amperes per seconds, or

2. 125,000 amperes for initial rates of rise of current above 5 million amperes per second.

For multisource d-c systems, semi-high-speed breakers can be applied in circuits where the current does not exceed 250,000 amperes in 0.25 second, but possible damage can occur to the breaker.

Alcoa's Rockdale system is composed of many power conversion units and like all other high-capacity multisource systems, the parallel combination of these units gives a low system time constant. Even with these high rates of rise, the short-circuit duty to the system does not exhibit the offset transient that appears in a single unit with a fault directly on its terminals. The physical spacing of the system plus the loss of components on overload protection tend to eliminate the extremely high calculated current levels.

The limits from the standard are illustrated in Fig. 8 for comparison with breaker performance. Here, the available current at 0.025 second, using the maximum short-circuit conditions, was approximately 220,000 amperes with an ultimate level more than 400,000 amperes. Even with these high levels, the breaker completed the interruption with normal arcing and limited the time to the crest current of 0.031 second. The standard has given 0.030 second as the time to the current crest when functioning in conjunction with available currents of 40,000 to 125,000 amperes at 0.025 second. The duty shown is much higher with the time to crest being only 1 millisecond longer.

The second test shown in Fig. 8 shows the available current held to a value under 125,000 amperes at 0.025 second, and the tripping means functioning to restrict the crest time during interruption to 0.014 second.

The interrupting ability of the d-c air circuit breaker, Fig. 4, in the tests has met the service requirements. Therefore, it would appear that the limits given in the standard for semi-high-speed applications are too conservative in relation to the capability of semi-high-speed d-c air circuit breakers.

Conclusions

This extensive testing program on Alcoa's system has provided information about the system and the performance of semi-high-speed d-c air circuit breakers. These conclusions can be made:

1. Semi-high-speed d-c breakers can adequately interrupt the extreme short-circuit current levels of a system, calculated with an available short-circuit capacity of 880,000 amperes at the bus and measured at the test point to be more than 400,000 amperes.

2. Multisource d-c systems are usually of higher rate of rise than associated with the standard for the semi-high-speed d-c breakers.

3. With delayed removal of a bus short circuit, the usual multisource system will separate and drop out units on overload.

4. Additional consideration should be given by the industry to the basis for interrupting performance and rating of semi-high-speed and high-speed d-c air circuit breakers.

References

1. Large Air Circuit Breakers. *NEMA Standard SG3-1955*, National Electrical Manufacturers Association, New York, N. Y., 1955.

2. A Mathematical Model and Procedure for Arc-Back Current Calculations for Power Rectifiers, H. P. Fullerton, J. Teno. *AIEE Transactions*, vol. 77, pt. II, Nov. 1958, pp. 456–64.

3. Comparison of Calculated and Measured Arc-Back Currents in Large Power Rectifier Systems, J. Teno, C. H. Titus, R. N. Wagner. *Ibid.* (Jan. 1959 section), pp. 585–89.

4. Short Circuit Calculating Procedure for Direct Current Systems, AIEE Working Group. *AIEE CP 57-811*, sect. X (available on request).

Discussion

C. I. Clausing (I-T-E Circuit Breaker Company, Philadelphia, Pa.): The tests described in this paper represent a considerable addition to the available knowledge about high-capacity d-c system faults. The authors are to be commended for making the most of the opportunity to run a very good and comprehensive set of tests.

We find it necessary, however, to take exception to conclusion 1 drawn from these tests. This conclusion states that semi-high-speed d-c breakers can adequately interrupt the extreme short-circuit currents of this type of system. From a circuit-breaker standpoint, this is true, but a prerequisite is that the system, to be protected, must also be adequate to withstand the mechanical and thermal stresses involved. In many cases the extra bus bracing, higher quality bus insulation, or increased copper section would be hard to justify economically.

The reason for this extra bracing in this case is that the peak value of let-through current is 220,000 amperes in 0.031 second whereas the standard for high-speed d-c breakers would limit this to a peak of approximately 104,000 amperes in 0.008 second.

We believe that the limitations on breaker interrupting ratings set up under reference 4 of the paper represent the practical considerations for getting a maximum of system protection.

It has been our experience that a substitution of high speed for semi-high-speed protection results in a more economical, practical system for the user.

P. L. Hartsock, T. J. Scully, and V. N. Stewart: The authors appreciate the discussion of Mr. Clausing and its support of the value gained by this over-all test program. Through these tests a knowledge of both the system characteristics and breaker performance was obtained. Both these items are very valuable to the application engineer who is frequently faced with problems on systems which exceed the limits given by breaker standards.

The first conclusion giving performance of the breaker at available currents of 400,000 amperes was positively demonstrated in the program. This was a semi-high-speed breaker and represents a type of breaker having wide application in the industry where large quantities of d-c power are used. The past performance of these systems indicate the adequacy of the normal bus designs versus the interrupting ability. The degree by which economies can be made in the high-capacity systems by the use of high-speed breakers would still have to be evaluated for each project. It is not expected that the paper eliminates consideration of high-speed breakers by the application engineer. Rather, it has given him further information on the interrupting duties which he combines with the system design and operation plan to achieve the necessary protection and desired operating conditions.

The facts presented in the paper should also aid in the future review of standards and guides so that the adopted limits will provide maximum system flexibility and adequate protection.

Freight Car Tractive Resistance Measurements by Doppler Radar

R. D. CAMPBELL
NONMEMBER AIEE

MODERN classification yards permit an incoming freight train to be redistributed to tracks assigned to various destinations, with all measurements, computations, and speed control performed automatically. In such yards cars are uncoupled at the crest of a hill and then accelerate freely down a slope which is usually about 3%. This increase in speed separates the cars so that track switches may be safely thrown between cars. Some 250 feet down the hill is the master retarder, a pneumatic or electric brake, which squeezes the wheels of each car between brake beams and reduces the car speed in a controlled manner to a value which can be handled by subsequent retarders. Below the master retarder the tracks fan out to a number of group retarders; here the exit speed must be computed which will allow each car just sufficient energy to traverse the subsequent curve and travel the free roll distance down straight classification track to a gentle coupling with cars already there. A radar-controlled servo loop is closed around each retarder so that the desired exit speed is accurately attained.

Two independent measurements are made on each car to determine separately the energy losses on straight and curved track. The curved track computation is based on two velocity measurements obtained by radar on the stretch of track between master and group retarders. During yard installation a number of test cars are run into each track to determine the correlation between this measurement and energy loss in the curves below the group retarders, and these correlations are permanently set into the computer by potentiometers. The correlations so obtained are usually not very good, since car behavior on curved track depends on many more physical properties of the car than does straight track behavior. Probable errors of ±25% are not uncommon, but this precision is usually adequate, since yard curves are rarely more than 500 feet long. The straight portion of each classification track, however, may be several thousand feet long, and accordingly the prediction of energy loss per unit length of straight track must be considerably more precise.

This paper deals with the measurement of energy loss on straight track. The quantity actually determined is the deceleration which the car would experience on zero grade and therefore includes the components of friction between rail and wheel, within the bearings, and the accelerating or decelerating effect of wind. This deceleration is often called rolling resistance and its symbol used throughout this paper is R_t. It is a quantity like a coefficient of friction, so that the whole decelerating force on a car is obtained by multiplying R_t by the car weight. In typical cars these coefficients are very small, of the order of 0.001 to 0.003. For practical computation it is more convenient to express R_t in pounds retarding force per ton of car weight therefore

Paper 58-1212, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 30, 1958; made available for printing August 22, 1958.

R. D. Campbell is with Union Switch & Signal, a Division of Westinghouse Air Brake Company, Swissvale, Pa.

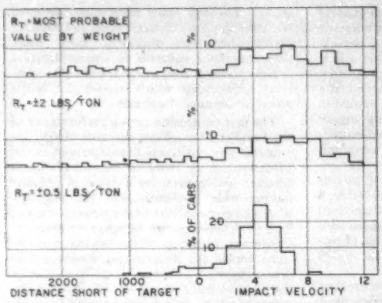

R_t = MOST PROBABLE VALUE BY WEIGHT
R_t ±2 LBS/TON
R_t ±0.5 LBS/TON
% OF CARS
2000 1000 DISTANCE SHORT OF TARGET
0 4 8 12 IMPACT VELOCITY

Fig. 1 (left). Computed automatic yard behavior

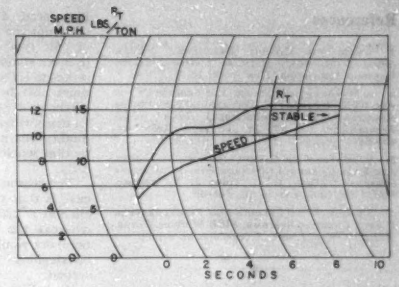


Fig. 2 (right). Stable light car R_t record at 26 F

yielding quantities 2,000 times as big. A typical 75-ton car is then said to have a value of R_t of about 3 pounds per ton. This means that the total force decelerating the car is 225 pounds, or that the coefficient of friction is 0.0015. One pound per ton therefore means a friction coefficient of 0.0005, a deceleration of 0.0005 g, or a deceleration of about 0.01 mph (miles per hour) per second. Since a measurement of R_t must have a probable error of the order of 0.5 pound per ton, it is evident that very small quantities are involved.

The requirement of a probable error in R_t no greater than ±0.5 pound per ton comes from a Monte Carlo analysis of the automatic classification yard problem. Fig. 1 shows the computed behavior of a yard when cars are expected to go 3,000 feet down a track whose accelerating grade is 2 pounds per ton. It is clear that using the most probable value of R_t for a particular car weight, or a measured value of R_t with probable error of ±2 pounds per ton, results in equally unsatisfactory behavior: many stalled cars and many couplings at speeds above 6 mph. Great improvement results, however, when R_t can be measured with a probable error of ±0.5 pound per ton.

Transient Effects in R_t

The measurement of R_t in an automatic yard is usually done while the car is freely accelerating down a grade which is typically about 3%. In general, when accelerating from slow speeds, cars display transient behavior, with R_t leveling off at a stable value for speeds above 10 mph. The initial transient is particularly severe in cold weather for empty cars, but cannot be ignored for any weight car in any weather. There are also some cars which never display a stable value during the speed range of 4 to 12 mph. To date, we know of no way of utilizing data from these latter cars to predict their behavior in the classification tracks.

Fig. 2 is a composite of several car records made at different locations on the slope to illustrate the behavior of a typical empty car in cold weather. With this car a stable value was attained at 9.6 mph. Fig. 3 is a similar composite record of a car which is never stable through the measuring range. Fig. 4 is a shorter portion of a record for a tank car, showing the 4-second period of liquid motion which is typical of tank cars that are not quite full.

Fig. 3 is not to be considered typical of transient R_t behavior; in fact, there are as yet no useful generalizations about the shapes of these transients. Cars have been observed whose rolling resistance continues to increase as they accelerate down the hump; others display initially a large value of R_t which continuously decreases. Still others may show one or two plateaus or a maximum or minimum. There is some experimental evidence that the later behavior of cars in the classification tracks may show a velocity dependence:

$$R_t = K_0 + K_1 v + K_2 v^2 + \ldots .$$

However, on account of the great variations in transient behavior which cars may actually display during their first 250 feet of free roll, it appears at present impossible to compute such velocity dependence from any measurements which can be made on the slope of a typical hump.

On account of these various transient effects, it is evidently necessary not only to measure R_t but also the rate of change of R_t and accept as valid only those measurements for which the rate of change remains small for a long enough time to rejects cases like Figs. 3 and 4. This

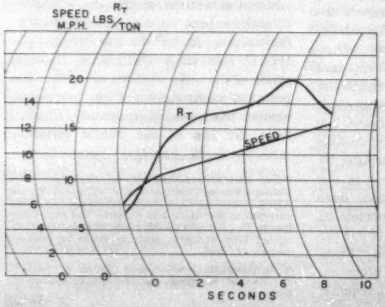


Fig. 3 (left). Unstable light car R_t record at 22 F

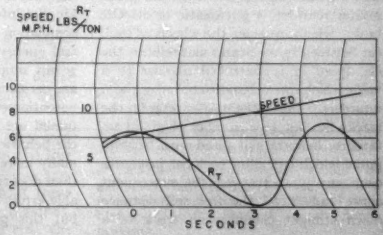


Fig. 4 (right). Medium tank car R_t record

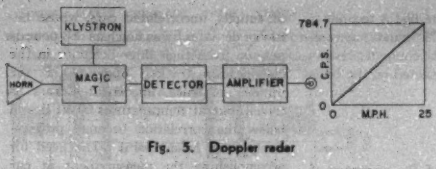


Fig. 5. Doppler radar

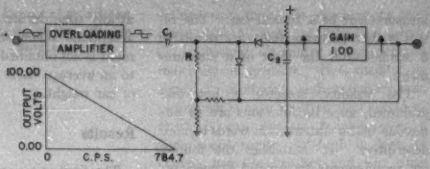


Fig. 6. Frequency meter

necessity rules out some easier techniques which might otherwise be used. For example most cars cross the apex of the hump at 2 mph, and therefore their arrival speed, some 250 feet down the slope, is given a measure of an average value of R_t. This value, however, is greatly contaminated by the initial transient, and yields no information upon which to base a validity decision. A second method, using two travel times, can be applied far enough down the slope to avoid most of the initial transient trouble. However, in this case also the computer output is a single numerical value, from which stability of R_t cannot be inferred. For these reasons, a very thorough research has been conducted with Doppler radar, to obtain velocity records sufficiently accurate that a first derivative can be precisely computed, and that a second derivative can be obtained which is adequate for validity decisions.

The Measuring Platform

A very necessary part of the instrumentation is the measuring track itself. It need be only 100 feet long, but accurate in grade, cross level, and gage. Since the instrumentation measures car acceleration, it is necessary to subtract this acceleration from track grade to get R_t. Accordingly an uncertainty of 0.5 pound per ton in the grade makes an uncertainty of 0.5 pound per ton in R_t. Adequate precision of grade can be attained by carefully shimming the track, the thinnest shims used being 1/8 inch. Long-time stability of grade is, of course, very desirable; one way of attaining it is to mount the track on a reinforced concrete slab well supported by driven piles. In some yards a scale track is available for enough downhill that cars reach the 10-mph minimum speed necessary for most of them to display steady-state behavior. Such scale tracks require the same order of precision for accurate weighing, so that R_t measurement can be accomplished without special track work. In other yards the scale is too far uphill, or nonexistent, and then an accurate section of track must be included in yard design solely for R_t use.

In a new yard, there will inevitably be some settling of track structure even when mounted on concrete. It is, therefore, desirable to schedule a frequent survey with elevations measured at 10-foot intervals along the track. From these data, the elevation differences are computed for every pair of points 40 feet apart, which is approximately one car length. When irregular settling results in a total spread of elevation differences greater than 1/4 inch, reshimming is indicated.

Most of the experimental work reported in this paper was accomplished at the Chesapeake & Ohio Manifest Yard at Russell, Ky. It was there established that a track crew and surveyor operating as a team, could shim a section of track with the needed precision, and without any unusual equipment. The grade of the Russell track, when averaged over each car length and further averaged by the smoothing time of the differentiator, is specifiable to ±0.1 pound per ton. As a contrast, the grade of 100 feet of track laid entirely on rock ballast in another yard, and power tamped in an effort to acquire accurate grade turned out to be specifiable to ±1.4 pounds per ton.

The Instrumentation

Fig. 5 is a block diagram of the Doppler radar. At the detector a portion of the outgoing radar signal is combined with the reflected signal from the moving car to yield the Doppler difference frequency which is directly proportional to target speed. The transfer function is also shown; with radar frequency of 10.525 kilomegacycles per second, the Doppler frequency is 31.388 cps (cycles per second) for 1-mph target speed. Therefore, 25-mph car speed produces a Doppler frequency of 784.7 cps. The antenna is a horn, polarized with E horizontal, mounted outside the rails, with the top slightly below rail head height. With this geometry, radar fadeouts due to multipath cancellations occur on only 1.5% of the cars. Other antenna arrangements can easily yield fadeouts on 30% of the cars.

Fig. 6 represents the frequency meter circuitry, which produces a slowly changing voltage proportional to a slowly changing Doppler frequency. The technique is fundamentally the dipper-bucket frequency meter with an amplifier whose gain is accurately 1.00 providing the bus to which the diode clamps the dipper capacitor C_1 during charging. The large resistor R provides low-frequency compensation by a nonlinear method, to balance the effect of a finite bucket capacitor C_s. The over-all transfer function of the frequency meter is linear within 1 part in 5,000.

Fig. 7 is the block diagram of the differentiator. R and C determine the properties as a differentiator, and the other components provide filtering for unwanted high-frequency accelerations which are actually present in the motion of the car. At low frequencies the output is directly proportional to frequency, and at high frequencies the output drops off proportional to the cube of the frequency. The turnover is at 0.25 cps. Not shown is relay circuitry which during stand-by maintains charge on the capacitors near the final values which they will attain on

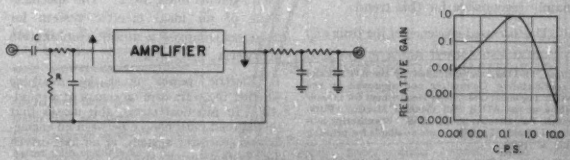


Fig. 7. Differentiator

measurement of a typical car. The result is that accurate data are available 3 seconds after the start of a measurement.

The validity detector is not diagrammed, since its functions are so numerous that a statement in words is more descriptive. It establishes the following requirements for a valid R_t measurement:

1. Three seconds must elapse after the start of a measurement, to assure the end of the instrumentation transient.
2. The rate of change of R_t must be less than ± 0.4 pound per ton per second.
3. The value of R_t must lie between 0 and 25 pounds per ton.
4. Conditions 2 and 3 must persist for 0.7 second after condition 1 has been fulfilled.
5. Condition 4 must occur before the first wheel of the car leaves the measuring section of track.

In typical yard geometry, at least 6 seconds are available for observations on each car. Therefore, about half the interval is devoted to the response time of the instrumentation and a valid measurement lasting 0.7 second and occurring at any time during the last 3 seconds is accepted. Whenever a valid measurement is not obtained the computer resorts to an average value selected on the basis of car weight.

Results

The final criterion of the adequacy of this instrumentation is the correlation between measurements so obtained and measurements on the same cars over long lengths of track in the classification yard. The latter measurements can be made accurately enough during tests by stationing two observers 500 feet apart and measuring travel times through 39-foot rail lengths by stop watches. Such measurements at Russell, Ky., produce a Gaussian distribution of points around the median with a probable error of ± 0.6 pound per ton, thus demonstrating that the instrumentation is free of systematic errors. The temperature range covered by data so far obtained extends down to 40 F (degrees Fahrenheit).

Similar measuements on the uphill scale track at Russell show a large number of entirely uncorrelated cars, whose behavior deviates by as much as ± 3 pounds per ton from their later behavior in the classification tracks. Above 60 F the number of such uncorrelated cars is trivial, but at temperatures of 50 F and below, the correlation becomes progressively more unsuccessful. The need for accomplishing the measurement at car speeds of at least 10 mph is therefore experimentally verified.

Conclusions

The instrumentation described in this paper is now considered precise enough to be included in classification yard automation. Credit for the success of this research must go to the participating railroads whose facilities were used for experimental work. Important parts of the research were carried out at Conway Yards of the Pennsylvania Railroad and Gateway Yard of Pittsburgh and Lake Erie Railroad. The largest share of the experimental work, however, was acomplished at the Chesapeake and Ohio Manifest Yard at Russell, Ky.

A Novel Generating System for Railroad Cabooses

L. B. HADDAD
ASSOCIATE MEMBER AIEE

R. A. VERCELLA
ASSOCIATE MEMBER AIEE

D. W. BROWN
ASSOCIATE MEMBER AIEE

FOR MANY YEARS only an insignificant fraction of the cabooses operating on America's railroads have utilized any form of electric power. Most cabooses have been heated by oil stoves and illuminated by kerosene lanterns. Signaling and communications have been carried out by means of written notes or hand signals.

In recent years, however, there has been an increasing trend on the part of the railroads to use electric equipment in cabooses. Three factors have been mainly responsible for this trend:

1. The increased efficiency of the train and crew made possible by the use of electric communication equipment.
2. Consideration for the safety and comfort of the train crew.
3. Legislation requiring the use of electric power in cabooses in many states.

Paper 58-1159, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted March 13, 1958; made available for printing July 25, 1958.

L. B. HADDAD, R. A. VERCELLA, and D. W. BROWN are with Safety Industries, Inc., New Haven, Conn.

Requirements for Caboose Generating Systems

Increasing demand for electric systems on cabooses has led to careful consideration of the technical requirements which such systems must meet. The specifications of an ideal electric system for cabooses comprise a unique combination of criteria:

1. Electric power must be available whether the train is in motion or at a standstill. In practical terms, this means that the railroad must choose between an engine-driven generator system, or an axle-driven d-c system using storage batteries. Although batteries of reasonable size can sustain full electrical load for only a limited of time, most railroads prefer this restriction to the high maintenance and operating costs of an engine-driven generator.
2. The most popular system voltage for cabooses is 12 volts d-c. Most communication equipment suitable for use in a caboose is available with power supplies designed to accomodate it. In order to charge lead-acid batteries at an acceptable rate with a nominal 12-volt system, the output voltage of the generator should be set at 13.8 volts; with nickel-iron batteries the voltage setting should be 15.5 volts. Regulation of $\pm 3\%$ is required over the full range of speed, load, temperature.
3. The power rating of the generating system applied to the caboose will depend upon the amount of electric equipment which the caboose carries, and the amount of reserve power which must be available for charging the batteries. At present a nominal output rating of 1.5 kw is meeting the greatest acceptance.
4. The speed range over which power is required is very broad. A caboose attached to a short train of emtpy cars may travel as fast as 80 miles an hour, whereas a heavily loaded train on a sustained upgrade may move at less than 10 miles an hour. The electric system must be able to maintain regulated voltage and carry load under both of these conditions.
5. Equipment attached to the underframe of a caboose must endure the most severe conditions of high shock, heat, cold rain, snow, ice, flying rock ballast, and wind-driven dust. Conditions inside the caboose are somewhat less severe, but still not com-

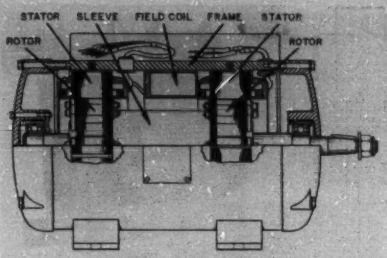


Fig. 1. Layout of an inductor alternator design for caboose service

parable to those which would be expected inside railroad passenger cars, or in most industrial plants.

6. Because the railroad personnel who presently maintain caboose equipment have little or no familiarity with electric systems, it is of the utmost importance that a caboose electric system operate with an absolute minimum of maintenance.

7. Railroads demand equipment life which is long compared to that usually required by industrial or military users. The generator, for example, should last for over 25 years with only bearing replacement. Rectifiers and regulators should operate for well over 10,000 hours with only a minimum of attention.

8. The selling price of the caboose electric system must be as low as possible consistent with meeting the foregoing objectives.

After a study of these requirements, the Electrical Division of Safety Industries, Inc., undertook the design of a 15-volt 1.5-kw d-c generating system which would consist of a power transmission, an alternator, a rectifier, a battery, an adjustable voltage regulator, and means for overload protection. Certain components of this system involve design features never before offered in railroad service.

Alternator

Despite the fact that most caboose electric systems are d-c systems, very few caboose generators are d-c generators. The reasons for this are found in the requirements for a caboose generating system. Since the operating speed range of the generator must be very wide, and the maximum speed of a d-c generator is limited by commutation and mechanical considerations, the cut-in and minimum full-load speed of a d-c caboose generator will be extremely low. Furthermore, the requirements for reliable operation through severe climate conditions, coupled with the fact that the d-c machine has a commutator and brushes to be protected, mean that the d-c generator ought to be totally enclosed. The resulting machine will be almost prohibitively large, heavy, and expensive.

The majority of generators operating on cabooses up to the present have been alternator-rectifier systems, and the alternators have all been of the monocoil type. (A monocoil alternator may be defined as a multipole synchronous machine, with a single rotating field winding.) Such an alternator is compact and relatively inexpensive. It is readily designed to produce a high-frequency output which is easy to filter after rectification. It can operate reliably over a wide speed range, although it does, of course, use the brushes and slip rings which characterize all rotating field alternators.

In an effort to approach the ultimate in ruggedness and reliability, Safety Industries has discarded the synchronous design in favor of an inductor alternator. The inductor alternator offers mechanical simplicity comparable to that of a permanent-magnet alternator, while retaining the use of a controllable field winding. Since the field winding is located in the stationary member of the machine, brushes and slip rings are unnecessary. The rotating member, in fact, contains no windings whatsoever, making this type of machine especially well adapted for very-high-speed operation. Like the monocoil synchronous generator, the inductor alternator lends itself to multipolar configuration. Even at low shaft speeds it can produce a high-frequency output which is easy to filter after rectification. In addition, the inductor alternator is readily designed so as to produce a "square" waveshape, which tends to reduce the amount of ripple in the rectified voltage.

For those not familiar with inductor alternators, an explanation of the operation of the machine may be in order. The layout of the inductor alternator design actually used in this caboose application is

Fig. 2. Interior of an inductor alternator for caboose service with the bearing bracket removed

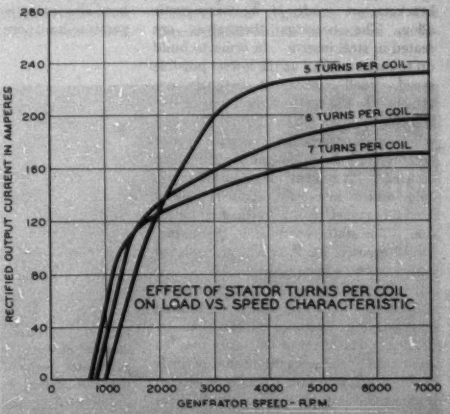


Fig. 3. Relationship between maximum power capacity, generator speed, and turns per stator coil, for identical inductor alternators with regulated voltage of 13.8 volts

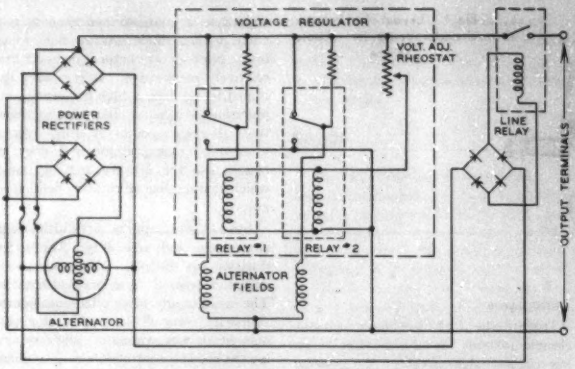


Fig. 4. Schematic diagram of a complete generating system, including alternator, rectifiers, line relay, and voltage regulator

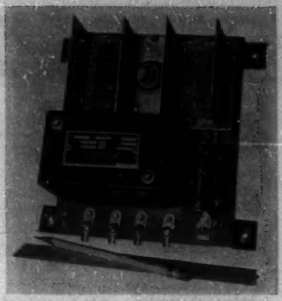


Fig. 5. Experimental voltage regulator, utilizing a transistor

shown in Fig. 1. It will be seen that the field winding encircles the shaft and sleeve. The flux path, therefore, passes through the shaft and sleeve, outward through one of the rotors, through the stator, back through the frame to the other stator, and inward through the rotor to the shaft and sleeve again, thus completing the magnetic circuit. The rotors are design with one tooth for every two stator poles. When a tooth is beneath a stator pole, the winding in that pole links maximum flux. When a tooth is not beneath the pole, the flux falls practically to zero. This varying flux, linking the stator winding, induces an alternating voltage.

To prevent the bearing brackets from becoming a leakage flux path, the bearing brackets are made of a nonmagnetic alloy. The bearings themselves are seated in steel inserts. In order to build up to rated voltage at the lowest possible speed, small permanent magnets are attached to the teeth of the rotor of this machine; Fig. 2. Were the generator not self-excited, or were a higher cut-in speed acceptable, these magnets could be omitted. Permanent magnets, however, offer added insurance against loss of residual magnetism due to shock or vibration. The stator windings in this inductor alternator have been encapsulated with a thermosetting resin.

For an inductor alternator of a given size, maximum power capacity is limited primarily by reactance rather than by heating. The effect of reactance may be minimized by reducing the number of turns in the stator winding. This relationship is demonstrated in Fig. 3, which plots maximum rectified output current as a function of shaft speed for machines having identical magnetic circuits and a regulated output voltage of 13.8 volts. Note that although the machine with fewer turns per coil has the highest capacity at high speed, its output compares unfavorably at speeds less than 2,000 rpm. Since the maximum speed of all machines will be subject to the same mechanical limitations, the generator with the highest capacity will furnish power over the narrowest range of train speeds. Because of the very wide speed range demand in caboose service, a design with less than maximum capacity is used as standard.

Using a winding designed to give 108 amperes rectified output at 1,400 rpm heat dissipation is such that class *B* temperature limits are easily met with totally enclosed construction. This arrangement is preferable in caboose service to ventilated construction and class *A* insulation, because of the added protection afforded by the enclosed frame. Were the machine designed for higher outputs at higher speeds, ventilation would be both possible and necessary.

For the design adopted as standard, each stator of the alternator contains a single-phase class-*B* 7-turn winding rated 65 amperes at 16 volts a-c. The two stators are displaced 90 electrical degrees from one another, thus acting together as a 2-phase alternator. Full-load rectified output of 108 amperes is delivered at 1,400 rpm; the cut-in speed is 700 rpm, and the maximum speed (limited only by the power transmission) is given as 7,500 rpm.

Regulation and Protection

The company has adopted as standard a Tirrill-type voltage regulator specially designed for this application. In order

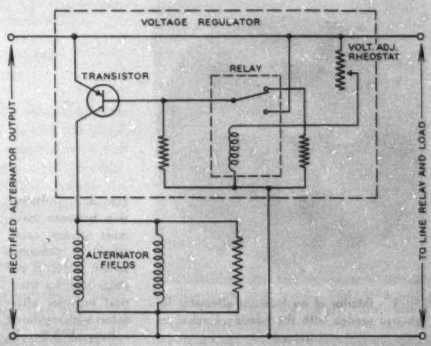


Fig. 6. Schematic diagram of the experimental voltage regulator shown in Fig. 5

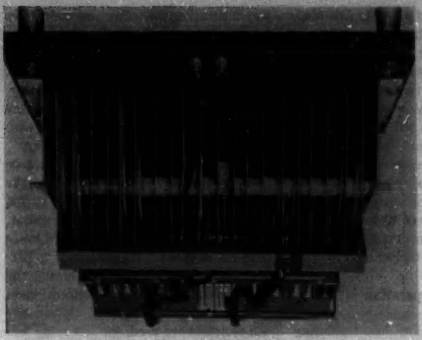


Fig. 7. Rectifier panel using selenium cells arranged for convection cooling. Plastic blocks containing the thermistor and thermostatic switch may be seen near the center of the rectifier in the foregound

to give acceptable contact life, two vibrating relays are used, each controlling one half of the field winding of the generator (Fig. 4). The resulting regulator combines low cost, reasonable life, very fast response, and close regulation.

In order to make the maintenance-free service life of the voltage regulator at least the equal of that of the generator and rectifiers, the company is now putting into service experimentally a new voltage regulator which utilizes a transistor. As shown in Fig. 5, the transistor is mounted on a finned heat sink, and the dust cover encloses a vibrating relay. The voltage-adjusting rheostat is mounted in the lower right corner of the panel. Other resistors are mounted on the back of the phenolic panel. In this new regulator, the transistor, employed in a common emitter arrangement, directly modulates the field current of the generator, and a vibrating relay controls the transistor; see Fig. 6. Through this arrangment, the relay switches only the base current of the transistor. Maximum power per contact of the relay is about 0.05 watt, and practically indefinite service life may be expected. Alternator field windings as shown in Fig. 6 are identical to those depicted in Fig. 4, although the windings are now connected in parallel. A resistor has also been connected in parallel with the field winding, in order to dissipate voltage surges when the field current is varied suddenly.

In any constant-voltage system using batteries, it is necessary to protect the generating system against overloads which can occur whenever it is asked to charge a highly discharged battery. This requirement is ordinarily met by furnishing a regulating device which will sense the amount of current being drawn from the the generator, and reduce the generated voltage to a point where a specified current is not exceeded. This arrangement is based on the assumption that this specified amount of current is all that can be handled without damage due to excessive temperatures.

In actual practice, the relationship between current and temperature is not absolute. While it is generally true that there is a maximum safe temperature for each part of the electric system, the amount of current necessary to produce this temperature depends upon the ambient temperature, upon the ventilation or heat-rejection conditions existing in the machine, and upon the length of time for which the current is drawn. A more appropriate type of power-regulating system, therefore, is one which measures temperature directly rather than current.

In this generating system, it was determined that the element most sensitive to high temperature was not the generator but the rectifier. The temperature of the rectifier will increase rapidly when overload is applied, whereas that of the generator will increase more slowly. Furthermore, the geneator is operating well below its maximum temperature rating, and can stand heavy overloads for appreciable periods of time without damage.

In the Safety system, the temperature of the rectifier is sensed by a thermostatic switch and a thermistor, both of which are embedded in plastic, and fastened in a tight fit between the plates of the rectifier; see Fig. 7. Until a temperature somewhat below the maximum safe temperature of the rectifier is reached, the thermostatic switch is open, and these two elements are disconnected from the regulator circuit (Fig. 8). As the temperature increases, the thermostatic switch closes and places the thermistor in parallel with the voltage-adjusting resistor which is in series with the voltage coils of the regulating relays. The thermostatic switch is used as an in-

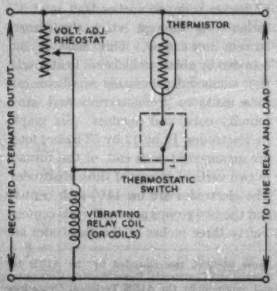


Fig. 8. Schematic diagram showing the application of the thermistor and thermostatic switch to the regulator circuit for purposes of power control

Fig. 9. Inductor alternator applied to the underframe of a railroad car

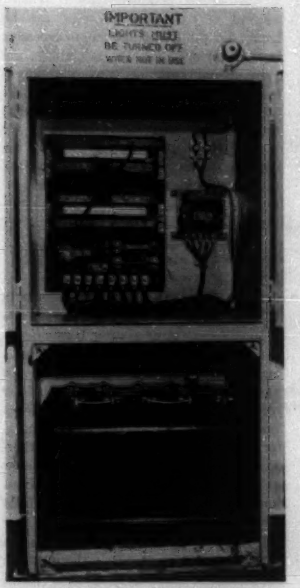


Fig. 10. Rectifier panel, voltage regulator, and battery, which have been arranged in a locker mounted inside a railroad car

expensive means of inserting the thermistor at a precise temperature, since thermistors made to very close tolerances are not readily available. Placing the thermistor in the circuit tends to increase the voltage coil current, lowering the regulated voltage, and partially unloading the system. Since the thermistor has a negative temperature coefficient, further increases in the temperature of the rectifier will bring about further decreases in regulated voltage, unloading the system further. This action will continue until the rectifier temperature stabilizes.

The effect of this power-regulating system is to permit small continuous overloads at ambient temperatures lower than the rated temperature of 40 degrees centigrade, and to permit moderate instantaneous overloads at any temperature.

Application and Operation

The generator, rectifier, and regulating systems may be applied to the caboose in a variety of ways. Because of the extreme ruggedness of the apparatus, little consideration need be given to protecting it, and the preferred arrangement will be that which is simplest and cheapest for the caboose in question. The generator is usually suspended from the underframe of the caboose, and is driven from the axle by a flat-and-V-belt drive, a 2-stage V-belt drive, or a gear transmission. If a gear transmission is used, it may be arranged with the propeller shaft either horizontal (Fig. 9), or vertical. In the former case, the alternator is driven through a horizontal propeller shaft from a gear transmission which is mounted on the end of one of the car axles. The rectifier and regulating system may be located either within the caboose cabin (Fig. 10), or in a ventilated enclosure attached to the underframe. In the former case, the rectifiers are selenium type, arranged for convection cooling. As may be seen from the illustration, line relay and main fuses are directly beneath the rectifiers, and a voltage regulator of the dual relay type is immediately to the right of the rectifiers. In the case of the ventilated enclosure attached to the underframe, the rectifiers and relay contacts are furnished with special covers against dirt and moisture.

When selenium power rectifiers are used, a line relay is incorporated in the system. The relay is responsive to the rectified output voltage of the alternator, and operates at an alternator speed lower than cut-in speed. Its purpose is to prevent discharge of the batteries through the reverse resistance of the rectifiers when the caboose is standing idle. When silicon rectifiers are used, the line relay will generally be unnecessary.

In a typical application, the generator will achieve cut-in speed of 700 rpm when the caboose is traveling at approximately 8.5 mph. At this point, the generator can maintain rated rectified voltage at no load. The ability of the generator to carry load increases as train speed increases, and rated load can be applied at rated voltage with a train speed of 17 mph. Regulated voltage is then maintained at all loads up to rated load, and all speeds up to the rated maximum, which corresponds to a train speed of about 90 mph. Overloads may, of course, result in a reduction of regulated voltage until the rectifier temperature has stabilized at a safe value.

This complete system is now in operation on several major railroads, and is performing with complete reliability throughout. Applications on certain types of marine and military equipment are also foreseen.

Characteristics of an Electric Resistance Furnace Load

W. A. STELZER
MEMBER AIEE

IN 1899 Dr. E. G. Acheson discovered that if he overheated silicon carbide, the silicon was volatilized out and pure carbon or graphite remained. Later tests showed that graphite could be manufactured by heating petroleum coke to a high temperature [2,600 C (degrees centigrade)].

The process partially described herein (see Fig. 1) uses rods 8.7 inches in diameter by 75 inches long made of calcinated petroleum coke held together by pitch. These rods are stacked vertically so as to touch each other crosswise of the furnace and are separated 3 inches apart lengthwise of the furnace. A current-conducting pack of new and old graphitized foundry coke mixture covers the rods in between them, at the bottom and sides, and on top. This results in a mass 87 inches wide, 85 inches high, and 355 inches long through which the furnace current must flow. This mass is surrounded by about 18 inches of heat insulation composed principally of silicon carbide made in previous runs and sand, foundry coke, and sawdust. Six graphite electrodes, 12 by 12 by 53 inches long, are arranged at each end, of the furnace in two vertical groups of three electrodes. The electrodes are on 14 1/2-inch centers and the two groups are on 32-inch centers. Thirty-three inches of the electrodes are

Paper 58-1258, recommended by the AIEE Industrial and Commercial Power Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted June 19, 1958; made available for printing August 28, 1958.

W. A. STELZER is with The Dow Chemical Company, Midland, Mich.

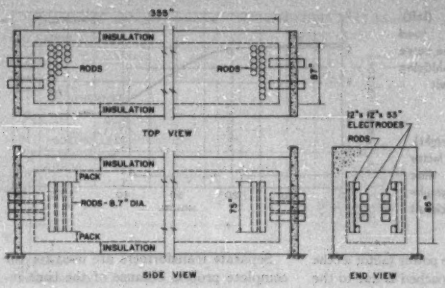


Fig. 1 (left). Graphite furnace assembly

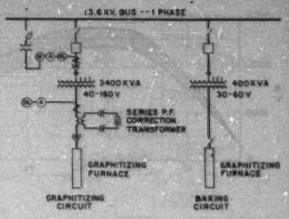


Fig. 2 (right). Diagram of graphitizer

in contact with the insulation and pack inside the furnace. The furnace current passes through the stacked mass perpendicular to the length of the rods.

The electrical characteristics of the furnace are greatly affected by the proportions of new foundry coke to old pack. Therefore these proportions must be held close to values determined by experience in order to use the graphitizing transformer equipment most effectively. The electric circuit for the furnace is shown in Fig. 2.

The first step in the graphitizing process is called baking. Current from a 400-kva (kilovolt-ampere) single-phase 60-cycle transformer is run through the furnace using an output voltage at the transformer of approximately 47 volts. The current in the furnace will gradually increase from about 1,100 or 1,300 amperes over a period of 408 hours to from 6,000 to 6,500 amperes; see Fig. 3 for a typical change in kilowatt load input over the period. Since the current increases with time, the transformer voltage must be reduced. This is done manually through a motor-driven under-load tap changer which is a part of the transformer. The conductors between the transformer and furnace are interlaced copper bus bar of 1/2-inch by 6-inch cross section. Suitable switches made of the same copper bar are located at the ends of the furnaces so that the baking current can be disconnected. The primary of the 400-kva transformer operates from a 13,800-volt source. The furnace temperature during this period rises from room temperature to 1,000 C. A typical curve for power-factor change during the baking period is shown in Fig. 4. This power factor is at the primary of the transformer.

After the baking period, the graphitizing transformer is connected to the furnace through water-cooled cables. The transformer has a capacity of 3,400 kva and operates single phase from the nominal 13,800-volt source, see Fig. 2. This transformer, together with its series capacitor transformer, is mounted on a standard-gage specially designed railroad truck with motor drive. The whole transformer assembly is moved into place at each graphitizing furnace (there are 16) on railroad rails. Specially designed jacks and plugs are provided on top of the transformer assembly for the connections to the 13.8-kv power source and the bank of askerel-filled capacitor units used in connection with the series capacitor transformer.

A typical run of the graphitizing transformer, which is the second step in the graphitizing process, begins with the voltage across the furnace terminals at 65–70 volts with a furnace current of about 17,000 amperes. The input to the furnace is held by means of manual under-load tap-changer control to a kilowatt input that will give a uniform temperature rise (about 40 C per hour) in the furnace with respect to time. This results in a gradual rise in the furnace voltage and current for a period of about 24 hours. At this time the maximum kilowatt that the graphitizing transformer can handle is reached. The kilowatt input is then held as high as possible but, due to the change in the power factor of the furnace, the top kilovolt-amperes rating of the graphitizing transformer becomes the maximum controlling factor and therefore the kilowatt input falls off some. At this part of the run the furnace voltage will be about 75–90 volts with the current at 35,000 to 40,000 amperes. After this time, which is about 40 hours of the run, the voltage of the furnace will stay fairly steady and the furnace current will increase somewhat but the kilowatt input will fall off a little (due to the graphitizing transformer primary current having to be held to under overload trip value) until about 60 hours of run is reached. The run is then considered finished and the power is shut off. Fig. 5 shows a typical kilowatt load versus time curve for the graphitizing run.

The typical graphitizing run can be expressed in terms of kilowatt and furnace resistance versus time. The kilowatt input is set at 400 kw at the beginning of the run and is uniformly increased to 2,300 to 2,700 kw in the next 22 hours. The power input is then held about con-

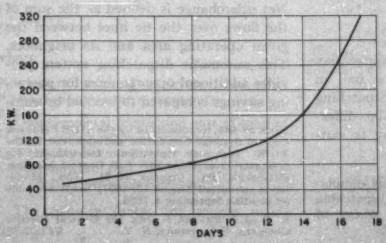


Fig. 3 (left). Kilowatt load versus time curve for baking circuit

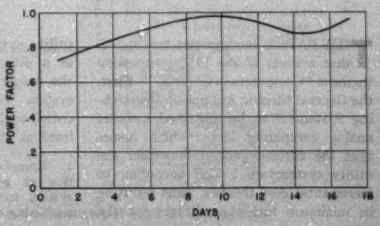


Fig. 4 (right). Power factor versus time curve for baking circuit

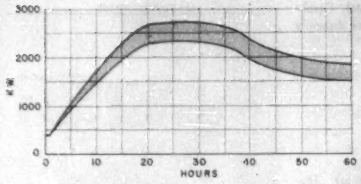


Fig. 5 (left). Kilowatt load versus time curve for graphitizing circuit

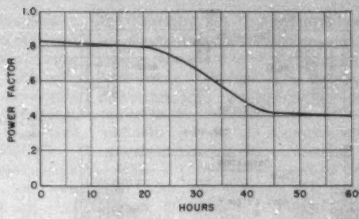


Fig. 6 (right). Power factor versus time curve for graphitizing circuit

stant for the next 10 to 12 hours at which time there is a tapering off of input to 1,400 to 1,900 kw at the end of the 60-hour run. During this time the resistance of the furnace decreases from about 4,700 microhms at about the 8th to 10th hour of the run to about 700 microhms at the 60th hour of the run. This change of resistance curve follows approximately that of a hyperbola having the equation $xy = +a^2/2$. Due to this change in resistance of the furnace, the power factor of the furnace varies approximately from 82% at the beginning of the run to near 40% at the end of the run. A typical power-factor curve for the graphitizing furnace is shown in Fig. 6.

Power-factor improvement as shown at the 13.8-kv input to the graphitizing transformer is accomplished by use of a series capacitor transformer and its capacitors connected in series with the furnace. This power factor varies approximately from 82% at the start of the run, remains nearly constant up to 24 hours of the run, and then gradually falls off to near 55% at the end, or 60th hour, of the run.

The falling off of the power factor as the end of the run is approached is due to the decrease in resistance of the furnace. Since the capacitor units in the series transformer secondary circuit are rated at 2,600 volts, the furnace current must be limited so that the 2,600-volt rating is not exceeded. If the series capacitor transformer and its capacitors could be run to a voltage approaching 5,000 volts, then the 82% power factor could be maintained. If the capacitor voltage could be increased as noted, then the other limitation would be the furnace current. The graphitizing transformer is not rated high enough to withstand the furnace current or the primary current needed to obtain the power-factor correction needed.

Since the graphitizing transformer equipment is not quite large enough to handle the furnace load, the alternative is to take the full 60 hours. It is believed that with large enough electric equipment, the graphitizing time could be reduced to 48 hours.

Separate transformers are used for the complete process because of the time required for the baking period. The graphitizing transformer can take care of four furnaces while the baking transformer is running on one furnace. In this process four baking transformers run simultaneously in order to keep the graphitizer occupied.

The use of under-load tap changers on these transformers was chosen for operating ease. Since the tap changer is operated frequently, its absence would entail excessive duty on the circuit breaker.

In conclusion, it is seen that the graphitizing furnace is a little too large for the graphitizing-period electric equipment. Also, as the electrical characteristics of the furnace load depend largely on the materials that are used to pack the furnace and the proportions used, the graphitizing transformer equipment can thus be used to best advantage.

(See Discussion on page 691)

Differential Analyzer Aids Design of Electrical Utility Automatic Dispatching System

L. K. KIRCHMAYER
MEMBER AIEE

THIS PAPER encompasses two important aspects of the 1957 symposium theme of "Computers in Control." First, the General Electric Automatic Dispatching System[1-2] to be described includes analog computing loops which insure that the turbine generators within the utility system are loaded according to incremental cost theories, thus resulting in minimum fuel expenditures for the utility system. Second, the application of a differential analyzer as a tool in the design of this automatic dispatching system will be described. The differential analyzer application may be outlined as follows:

1. Complete differential analyzer simulation of power system and dispatching system to determine optimum design of regulating and economic loops.

2. Differential analyzer simulation of power system tied together with prototype control components to study and evaluate the design of prototype equipment.

The General Electric Automatic Dispatching System simultaneously and automatically maintains for a given operating area, frequency, net interchange, and economic allocation of generation according to incremental theories. Net interchange is defined as the sum of the flows over the tie lines between the given operating area and its neighbors. This automatic dispatching system provides additional opportunities for operating savings compared to manual or semi-

Paper 57-946, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted August 12, 1957; made available for printing September 4, 1958.

L. K. KIRCHMAYER is with the General Electric Company, Schenectady, N. Y.

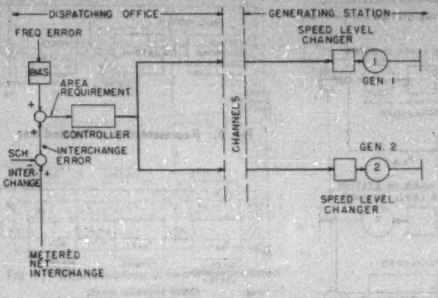


Fig. 1. Basic tie-line frequency control method

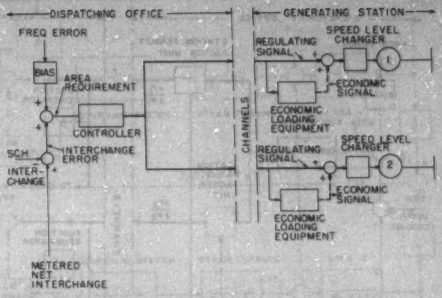


Fig. 2. Addition of economic loop to system of Fig. 1

automatic operation. These include the following:

1. Production economy. The dispatching system not only calculates the desired generator outputs for actual rather than forecasted conditions, but also automatically executes the schedule resulting in closer adherence to proper allocation of generation.

2. Manpower. Personnel previously employed in duties now performed by the dispatching system may be assigned other tasks.

Frequency and net interchange control is accomplished by means of a correcting signal developed from area requirement which is defined as the sum of the net interchange error plus a constant times the frequency error. The value of the constant determines the magnitude of the frequency bias. This correcting signal drives the speed level changer motors of the units within the area such that the area requirement is returned to zero.

Economic control is obtained by imposing additional control and computer devices on the basic tie-line frequency-bias control system. Optimum production economy for a given combination of machines in service is obtained when the incremental cost of received power is the same from all variable sources.[4] Stated mathematically

$$\frac{dF_n}{dP_n} L_n = \lambda \qquad (1)$$

where

F_n = input to source n, dollars per hour
P_n = output of source n, megawatts
$\frac{dF_n}{dP_n}$ = incremental production cost of source n, dollars per megawatt-hour
L_n = penalty factor of source n
λ = incremental cost of received power, dollars per megawatt-hour

The penalty factor for source n is defined as

$$L_n = \frac{1}{1 - \frac{\partial P_L}{\partial P_n}} \qquad (2)$$

where

$\frac{\partial P_L}{\partial P_n}$ = incremental transmission loss of source n

The incremental transmission loss[5] may be determined from a transmission loss formula as

$$\frac{\partial P_L}{\partial P_n} = \sum_m 2B_{mn}P_m + B_{no} \qquad (3)$$

B_{mn}, B_{no} = transmission loss formula coefficients

The choice of units to be operated is predicated upon spinning reserve requirements, area protection, voltage control, stability requirements, and economic considerations. The preceding co-ordination equation 1 determines the allocation of loading among the units chosen to be operated. Additional calculations are required to determine the most economic combination of units to place in service.

Description of General Electric Automatic Dispatching System

The fundamental components common to all tie-line frequency-bias controllers are indicated in Fig. 1. The area requirement signal is developed from the frequency and net interchange errors and is acted upon by a controller to develop a correcting signal to be sent to the speed level changer motors of all units under control. The control system causes the generator outputs within the area to take care of all load changes within the area.

Next superimpose upon the system of Fig. 1 an economic control loop as indicated in Fig. 2. By means of the economic controller, the actual unit output is compared to the desired unit output and

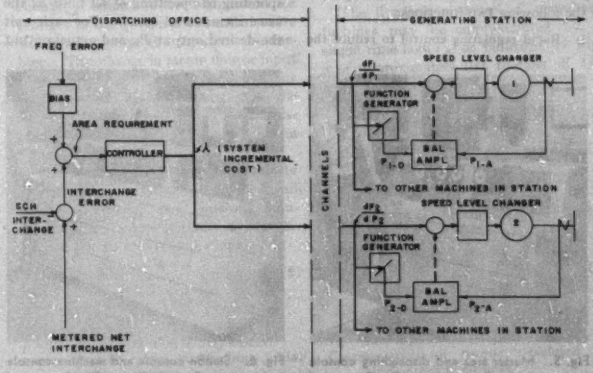


Fig. 3. System for automatic dispatching at equal incremental costs

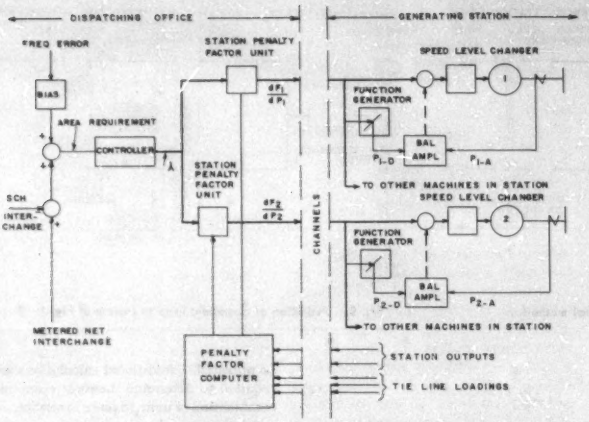


Fig. 4. Functional diagram of General Electric Automatic Dispatching System including penalty factor computation

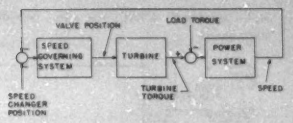


Fig. 7. Representation of isolated area

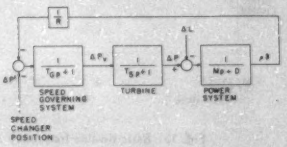


Fig. 8. Isolated area with illustrative transfer functions

a correcting signal is obtained which brings the actual output in agreement with the desired output.

First examine the operation of this system for the case in which transmission losses may be neglected. For optimum economy for this condition, the incremental cost of power of each generating unit should be equal to the same number λ. The co-ordination equations may then be written as

$$\frac{dF_n}{dP_n} = \lambda \tag{4}$$

In Fig. 3 interpret the signal generated by the controller as λ. This incremental cost is sent to all units and accomplishes the following two functions:

1. Rapid regulating control to reduce the area requirement signal to zero as indicated by the heavy solid line in Figs. 2 and 3.
2. Slower reset action to reallocate generation in the most economical manner corresponding to equation 4 as indicated by the dotted line in Figs. 2 and 3.

In addition to acting directly upon the speed level changer for each unit, the λ signal is fed into the incremental cost function generator for each unit. The function generator is an electronic data-storage device which indicates the relationship between the incremental production cost and output of the particular unit. For a given value of λ the function generators indicate the economic schedule corresponding to operating of all units at the same incremental cost λ. For each unit the desired output P_D and actual output P_A are compared and a correction signal is obtained. If the value of λ does not call for the proper total amount of generation, small changes in the area requirement signal occur which drive λ to its proper value.

Consider the case in which initially frequency and interchange are at their scheduled values and all units loaded at equal incremental cost. Next, assume that an increase of load occurs in the area concerned. An area requirement signal is immediately generated which calls for increased generation through the control indicated by the solid line in Fig. 3. As λ changes, these changes acting through the economic loops call for generation to be maintained at equal incremental cost. In the steady state after the load change has been absorbed by the units within the area, the area requirement signal is again zero and all units are operating at equal incremental cost.

Next consider the case in which transmission losses are significant. Optimum economy is then obtained when the incremental cost of received power is the same from all generating units. The form of the co-ordination equations is given by equation 1.

From equation 1 it is noted that the incremental cost of each unit is now modified by a penalty factor to determine the incremental cost of received power. Rewriting the equations as in the following will be helpful in understanding their application to the automatic dispatching system:

$$\frac{dF_n}{dP_n} = \frac{\lambda}{L_n}$$

Next, refer to Fig. 4 which indicates the introduction of penalty factor signals. Again λ is interpreted as the output of

Fig. 5. Master area and dispatching console in dispatcher's office of Kansas City Power and Light Company

Fig. 6. Station console and machine console in Hawthorn Station, Kansas City Power and Light Company

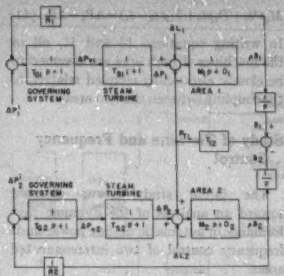


Fig. 9. Representation of two interconnected steam-electric areas

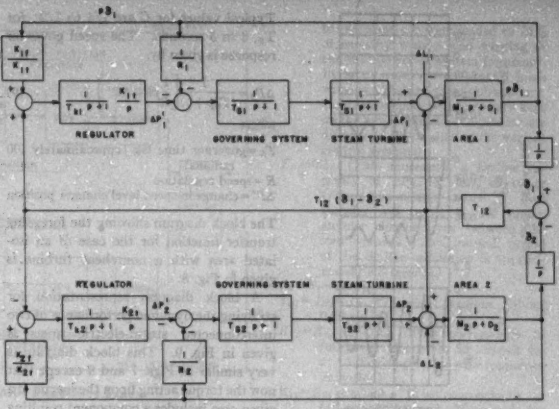


Fig. 10. Interconnected steam-electric areas with supplementary controllers

the controller. By modifying the λ signal for each station by the penalty factor as indicated, it is seen that the signal sent to each station will now be the proper incremental cost for that station. The penalty factor computer[4] requires for its inputs a knowledge of the station loads, interconnection flows, and a knowledge of the values of nonconforming loads. These quantities would most appropriately be telemetered information into the penalty factor computer. The $1/L_n$ signals are developed by an analog-computing circuit based upon the use of loss-formula coefficients as indicated by equation 3.

Fig. 5 shows the master area and dispatching consoles located at the load dispatcher's office. The master area console develops a signal proportional to the frequency and net power interchange error. This signal is interpreted as the incremental cost of received power. Means are provided in the dispatching console to modify manually or automatically this signal by the value of the plant's penalty factor, thus obtaining the appropriate incremental cost signal to be transmitted to each unit under control.

The master station console and the machine console such as that shown in Fig. 6 are both located at the station. The master station console provides all the adjustments for operating the whole station; the machine console provides adjustments for a given generating unit. In the machine console, the incremental cost signal from the dispatcher's office is fed into the incremental cost function generator. The function generator and servo-comparison loop insures that the unit is loaded in accordance with incremental cost theory.

As previously indicated, the master area console, which detects any deviation from desired conditions of system net interchange and frequency, adjusts the total generation to match the area load. This deviation is used to correct the prevailing incremental cost signal and restore the power system to balance.

Mathematical Model of a Power System

To study the dynamic performance of the automatic dispatching system, it is necessary to develop a mathematical model which represents the transient and steady-state performance of the power system. To assist in understanding the performance of interconnected areas, first examine the block diagram representation of an isolated area as shown in Fig. 7. The inputs to the speed-governing system are the speed signal and the speed level changer position. These signals call for changes in the governor-controlled valve which will change the input to the turbine. This change in steam flow or input to the turbine causes a change in turbine torque, which in conjunction with the load torque and the characteristics of the power system determine the change in system speed.

The following analysis pertains mostly to the response of the system for small changes. Writing the equation summing up the torques acting upon the inertia of the power system, yields

$$M\rho^2\delta + D\rho\delta = \Delta P - \Delta L \qquad (5)$$

where

M = effective rotary inertia of area
ρ = differential operation d/dt (t = time)
δ = deviation of rotary inertia from initial angular position
D = damping torque coefficient (i.e., the inherent net change of load and prime-mover torque with frequency)
ΔP = component of change in prime-mover torque produced by governor-controlled valve motion
ΔL = load change in area (aside from inherent change in frequency)
$\rho\delta$ = deviation from normal frequency or speed

All quantities will be expressed in per unit. Unless otherwise specified, time and angles are in radians. With 60 cycles per second as the normal frequency or speed, 1 second is equal to 377 radians.

Solving equation 5 for $\rho\delta$

$$\rho\delta = \frac{1}{M\rho + D}(\Delta P - \Delta L) \qquad (6)$$

This relationship is illustrated schematically in the block diagram of Fig. 8.

The dynamic response of nonreheat turbines may be approximated by a single time lag (T_s) as indicated by Fig. 8 of approximately 100 radians or 1/4 second. Here

$$\Delta P = \frac{1}{T_s\rho + 1}\Delta P_v \qquad (7)$$

where

ΔP_v = change in valve position

In the case of reheat turbines, the transfer function is given by

$$\Delta P = \left[\frac{CT_R\rho + 1}{T_R\rho + 1}\right]\left[\frac{1}{T_s\rho + 1}\right]\Delta P_v \qquad (8)$$

where

T_R = time lag associated with reheater
C = proportion of torque developed in high-pressure element

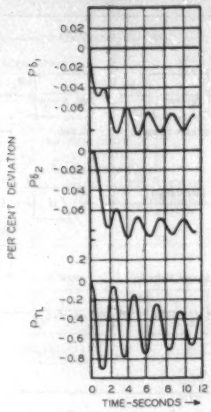


Fig. 11. Dynamic response with no supplemental control

The load change in Figs. 11, 12, and 13 is a unit function load disturbance of 1 per cent. The quantities $p\delta_1$, $p\delta_2$ and P_{TL} in these figures denote the frequency deviation in area 1, the frequency deviation in area 2, and the tie-line power deviation, respectively

ΔL in area 1
$R_1 = R_2 = 0.167$
$D_1 = D_2 = 0.75$
$T_{12} = 0.1$

Typical values for C are 1/4 to 1/2; for T_R, 3 to 5 seconds. The speed governor response is given by

$$\Delta P_v = \frac{-1}{(T_G p + 1)}\left[\frac{1}{R} p\delta + \Delta P'\right] \qquad (9)$$

where

T_G = governor time lag (approximately 100 radians)
R = speed regulation
$\Delta P'$ = change in speed level changer position

The block diagram showing the foregoing transfer function for the case of an isolated area with a nonreheat turbine is given in Fig. 8.

A block diagram representation for studying the dynamic performance of two interconnected steam-electric areas is given in Fig. 9. This block diagram is very similar to Figs. 7 and 8 except that now the torque acting upon the inertia of a given area includes a component resulting from the flow over the tie line. For example, the summation of torques acting upon area 1 is now given by:

$$M_1 p^2\delta_1 + D_1 p\delta_1 + T_{12}(\delta_1 - \delta_2) = \Delta P_1 - \Delta L_1 \quad (10)$$

where the subscripts 1 and 2 refer to areas 1 and 2, respectively, and

T_{12} = synchronizing torque coefficient of tie line
$T_{12}(\delta_1 - \delta_2)$ = change in flow over tie line = P_{TL}

Similarly, for area 2

$$M_2 p^2\delta_2 + D_2 p\delta_2 + T_{12}(\delta_2 - \delta_1) = \Delta P_2 - \Delta L_2 \quad (11)$$

In writing equations 10 and 11, all of the machines of each area are assumed to be closely electrically coupled relative to the coupling between the two areas.

Study of Tie-Line and Frequency Control

The initial study[7,8] was directed toward an analysis of the dynamic performance of automatic tie-line power-and-frequency control of two interconnected areas.

These analyses accomplished several specific objectives:

1. They confirmed theoretically (but modified slightly) the optimum ratios of frequency and tie-line power controlling effort previously based on reasoning from static conditions.

2. They obtained, in addition, optimum values of over-all control system gain which were not obtainable from the previous static analysis.

3. They showed the primary importance of close over-all system speed control which implies small governing-system dead band and the participation of all machines in speed regulation.

The dynamic performance of supplementary reset controllers was approximated by equations of the following form:

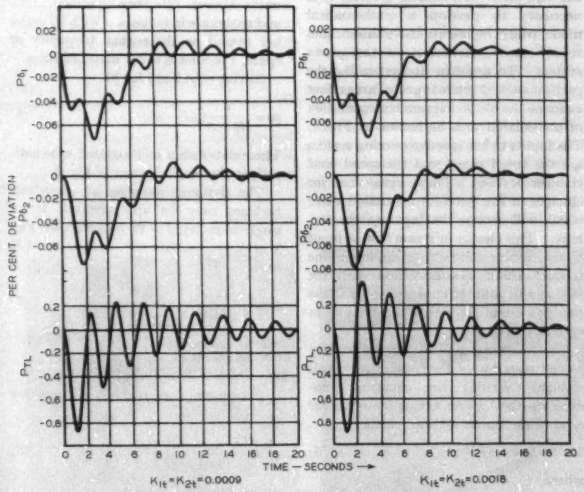


Fig. 12. Effect of tie-line power-controller gain

See subcaption, Fig. 11, with following addition:
$K_{1f} = K_{2f} = 0.0053$

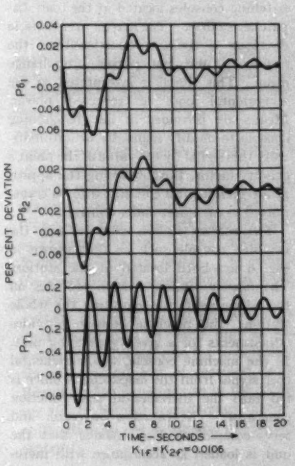


Fig. 13. Effect of frequency-controller gain

See subcaption, Fig. 11, with following additions:
$K_{1f} = K_{2f} = 0.0009$ and $K_{1f} = K_{2f} = 0.0106$

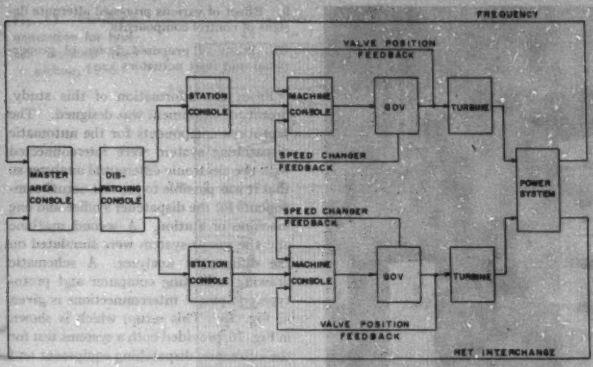


Fig. 14. Analyzer study of Automatic Dispatching System

$$p\Delta P_1' = \frac{K_{1f}}{T_{1s}p+1}\left[T_{12}(\delta_1-\delta_2)+\frac{K_{1f}}{K_{1t}}p\delta_1\right] \quad (12)$$

$$p\Delta P_2' = \frac{K_{2f}}{T_{2s}p+1}\left[T_{12}(\delta_2-\delta_1)+\frac{K_{2f}}{K_{2t}}p\delta_2\right] \quad (13)$$

where

K_{nt} = tie-line controller gain of area n
K_{nf} = frequency controller gain of area n
T_{ns} = time lag associated with regulator of area n

These equations represent the action of reset controllers since the rate of change of the speed changer position is proportional to the error signal. Consequently, the steady-state error given by

$$T_{12}(\delta_1-\delta_2)+\frac{K_{1f}}{K_{1t}}p\delta_1$$

must be zero, for example, for area 1. The preceding equations are based upon the assumptions that the time lags in the frequency signal and the tie-line signals are the same and that the feedback is from the synchronizing motor position. A block diagram representing the interconnected system with such supplementary controllers is given in Fig. 10.

Method of Analysis. The performance of the system as affected by variations in the parameters of the tie-line and frequency controllers was studied with the aid of a differential analyzer. The equations were set up on the analyzer and subjected to a unit-function load disturbance. The consequent variations of tie-line power, prime-mover powers, and system frequencies were plotted by the analyzer. Since an actual power system is subjected to random fluctuations of load, a set of response curves for load disturbances in each area in turn was obtained for each case. The values of the control system parameters resulting in the best compromise between rapidity and stability of response were considered to be the optimum. Specifically, a response having no overshoot was preferred to one more rapid, but with some overshoot. Thus, stability was emphasized over speed response. This choice of system parameters minimizes the continual oscillations produced by random load fluctuations.

System Parameters. The base system chosen for study had the following parameters:

$M_1 = M_2 = 4{,}000$ radians (corresponds to $H = 5.3$)
$D_1 = D_2 = 0.75$ per unit. (This was taken to represent a system loaded to 75% of capacity with a load varying as the first power of system frequency. In many cases power system loads vary as a higher power of frequency, but the damping was purposely kept on the low side in order to accentuate differences in stability of the various control systems)
$T_{12} = 0.1$. (This represents a tie-line capacity of 10%, or less, of either system capacity, as 10% load transfer will result in one radian angular displacement between the two systems. The natural period of the tie-line oscillation is approximately 2.3 seconds)
$R_1 = R_2 = 16.7\%$
$T_{G1} = T_{G2} = T_{R1} = T_{R2} = 100$ radians (about 0.26 second)
$K_{1f} = K_{2f} = 0.0053$. (This corresponds to full-power change in 30 seconds for a 1.0-cycle-per-second frequency deviation)
$K_{1t} = K_{2t} = 0.0009$. (This corresponds to full-power change in 30 seconds for a tie-line power deviation of 0.1 per unit.) Note that the base ratio $K_{nt}/K_{nf} = 17\%$. This has been a common value of effective speed regulation of power systems
$T_{S1} = T_{S2} = 100$ radians (about 0.26 second)

Results. The dynamic response of the system with no supplementary control is indicated in Fig. 11. Note that the overshoot on the tie-line approaches twice the final steady-state error. The period of the tie-line oscillation is determined primarily by the inertias of the two interconnected areas and the magnitude of the synchronizing torque coefficient.

Results illustrative of those obtained with supplementary control are given in Figs. 12 and 13. Fig. 12 shows the response of the system as affected by the tie-line power controller gain, and Fig. 13

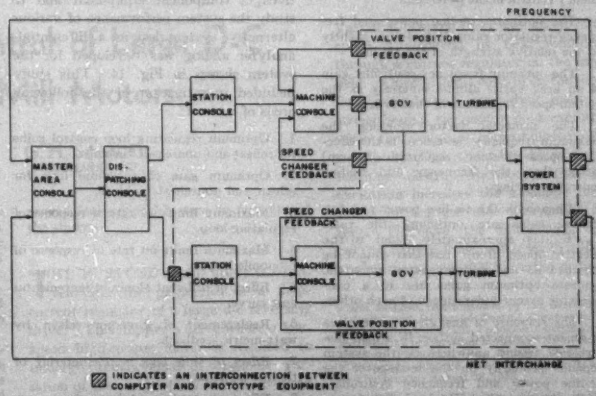


Fig. 15. Interconnection of prototype dispatching equipment with differential analyzer

Fig. 16. Development components of dispatching equipment interconnected with differential analyzer

similarly shows the effect of frequency controller gain. Both figures correspond to a system having an effective speed droop of 16.7%

From studies similar to this, the following conclusions were drawn:

1. Better, that is, more stable and rapid, response is obtainable with tie-line power-frequency bias control than with flat tie-line power control in one area and frequency control in the other area.

2. The optimum tie-line power and frequency-regulator gains depend only slightly on the relative strength of the tie line.

3. The optimum frequency-controller gain of an area varies almost inversely as the system-speed regulation of that area.

4. The principal factor affecting the maximum frequency deviations is the effective speed-governor regulation (droop) rather than the frequency and tie-line controller settings.

5. When both the tie-line power and frequency gains are optimum, the ratio K_{xt}/K_{xf} is approximately equal to the effective speed droop, but this cannot be regarded as an optimum ratio, since the separate optimum gains are to a considerable extent independent of each other.

6. The response of any given power area of an interconnected system to a load disturbance arising anywhere on the system depends primarily on the settings of the tie-line power and frequency controllers of the area in which the load disturbance originates.

7. In general, at values near the optimum, increasing the tie-line power controller gain increases the damping of oscillations following the initial swing, while increasing the frequency-controller gain decreases the damping of these oscillations.

Analyzer Study of Automatic Dispatching System

To determine optimum design objectives of component equipment and to study the system performance of various alternative system designs, a differential-analyzer analog was developed for the system shown in Fig. 14. This study included investigation of the following areas of interest:

1. Optimum regulating loop control gains for reheat and nonreheat turbines.

2. Optimum gain of economic loop for reheat and nonreheat turbines.

3. Maximum limits on rate of response of regulating loop.

4. Maximum limits on rate of response of economic loop.

5. Effect of different slopes of incremental cost curve.

6. Replacement of governor selsyn by watt-metric feedback.

7. Effect of time lags in telemetering of tie-line information.

8. Effect of dead band in various components.

9. Effect of various proposed alternate designs of control components.

10. Study of proposed design of proportional and reset actuators.

From the information of this study, prototype equipment was designed. The prototype components for the automatic dispatching system were interconnected with the electronic differential analyzer so that it was possible to provide actual components for the dispatcher's office and one machine or station. A second machine and the power system were simulated on the differential analyzer. A schematic drawing indicating computer and prototype equipment interconnections is given in Fig. 15. This setup, which is shown in Fig. 16, provided both a systems test for the automatic dispatching equipment and evaluation of the performance of the automatic dispatching system. This study included all of the following items:

1. Checkout of the electric system.

2. Confirmation of accuracy of previous computer studies.

3. Proportional and reset actuator performance.

4. Economic loading performance.

5. Response to step- and ramp-load changes.

6. Effect of backlash in various components.

Operating Experience

The automatic dispatching system decribed here has been in commercial operation since October 1, 1955, on the Kansas City Power and Light Company. Typical performance[3] is indicated in Fig. 17, which shows the generation on the Hawthorn units of the Kansas City Power and Light Company for a morning load pickup. Units 1 and 2 are represented with an identical incremental cost function. Units 3 and 4 have the same cost characteristic, this characteristic being slightly better than that for units 1 and 2. Initially, the load increases are picked up by units 3 and 4 while units 1 and 2 regulated about their lower limit settings. Later, load increases are shared by all machines when the incremental costs on all four machines become equal. As the incremental cost rises rapidly as full load is approached, units 3 and 4 take smaller proportions of the load increases after the total load has reached values corresponding approximately to 7 o'clock.

Conclusions

The General Electric Automatic Dispatching System has demonstrated

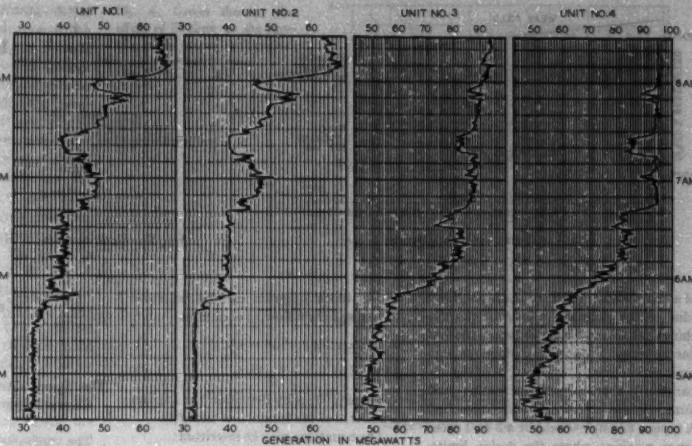


Fig. 17. Hawthorn generation for typical morning load pickup, 1956

through actual operating experience both the desired regulating and economic control and has resulted in significant annual savings.

The differential analyzer has been most helpful, both in examining the dynamic performance of the automatic dispatching system and in obtaining the optimum design of such a system.

References

1. How Modern Tools Improve Power System Operations, L. K. Kirchmayer. *General Electric Review*, Schenectady, N. Y., Sept. 1956.

2. A New Automatic Dispatching System for Electric Power Systems, K. N. Burnett, D. W. Halfhill, B. R. Shepard. *AIEE Transactions*, vol. 75, pt. III, Dec. 1956, pp. 1049–56.

3. Operating Experience with General Electric's Automatic Dispatching System at Kansas City, W. W. Scott, E. L. Mueller. *Minutes*, Missouri Valley Electric Association Engineering Conference, Kansas City, Mo., Apr. 18–20, 1956.

4. Evaluation of Methods of Co-ordinating Incremental Fuel Costs and Incremental Transmission Losses, L. K. Kirchmayer, G. W. Stagg. *AIEE Transactions*, vol. 71, pt. III, 1952, pp. 513–21.

5. Analysis of Total and Incremental Losses in Transmission Systems, L. K. Kirchmayer, G. W. Stagg. *Ibid.*, vol. 70, pt. II, 1951, pp. 1197–1205.

6. Penalty Factor Computers, H. H. Chamberlain, K. N. Burnett. *Minutes*, Annual Conference, Engineering and Operating Section, Pacific Coast Electric Association, Los Angeles, Calif., Mar. 21–22, 1957.

7. Tie-Line Power and Frequency Control of Electric Power Systems, C. Concordia, L. K. Kirchmayer. *AIEE Transactions*, vol. 72, pt. III, June 1953, pp. 562–72.

8. Tie-Line Power and Frequency Control of Electric Power Systems—Part II, C. Concordia, L. K. Kirchmayer. *Ibid.*, vol. 73, pt. III-A, Apr. 1954, pp. 133–41.

9. Effect of Speed Governor Dead Band on Tie-Line Power and Frequency Control Performance, C. Concordia, L. K. Kirchmayer, E. A. Szymanski. *Ibid.*, vol. 76, pt. III, Aug. 1956, pp. 429–35.

(See Discussion on page 692)

Motor Field Control of Large D-C Reversing Mill Motors

L. R. HULLS
MEMBER AIEE

G. H. SAMUEL
NONMEMBER AIEE

Synopsis: This paper analyses the performance of the field-current regulator on a large d-c reversing mill drive. A study is made of the transfer functions associated with the shunt-field-controlled d-c motor. The results obtained from an analog computer study of the problem are discussed and a brief description of the analog setup used for the study is given in Appendix I. Particular attention is paid to the performance of the current-limit control. The over-all study revealed certain weaknesses in the type of system used in practice, and the paper concludes with a brief description of a control system which embodies improved features which were an outcome of the study.

THE PURPOSE of this study is to investigate the properties of a field-current regulator of a large d-c reversing drive. This implies a drive where a low-speed high-torque motor is required to undergo rapid reversals of rotation. Variation of the motor speed from base-speed forward to base-speed reverse is achieved by control of the armature supply voltage. Variation of speed from base speed to top speed is achieved by field weakening through a current regulator on the motor field. Current-limit control on the motor field provides an overriding means of strengthening the motor field excitation when the armature current exceeds certain limiting values set by the particular machine design. Increasing the field excitation increases the available motor torque for a given value of armature current. Therefore, the motor with field-control current-limit, when set to operate above base speed, will have the property

Paper 58-1265, recommended by the AIEE Metal Industry and Industrial Control Committees and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted March 5, 1958; made available for printing August 28, 1958.

L. R. Hulls and G. H. Samuel are with the Canadian Westinghouse Company, Ltd., Hamilton, Ont., Canada.

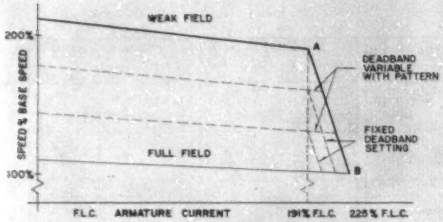


Fig. 1. Current-limit characteristic

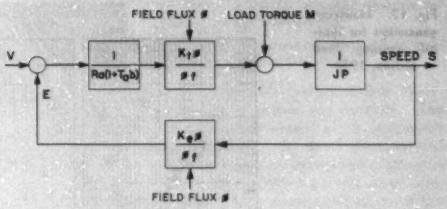


Fig. 2. Block diagram of d-c shunt motor

of a drive whose speed gradually decreases (due to the normal IR drop of the machine) as load increases: and when the load has increased to the point where current-limit action starts, the speed will fall comparatively sharply, with increasing load, until full-field excitation has been reached. A typical example would be a 2-to-1 speed range machine where the design limitations are 191% of full-load current at weak field and 225% of full-load current at full field. The design of the motor is such that the speed change, produced by a torque load which changes the armature current from 191% to 225%, should be a linear function of armature current. This is illustrated in Fig. 1.

A preliminary computer investigation of the problem showed that the current-limit characteristic obtained from a typical system was by no means a straight line between 191% and 225% full-load current (see Fig. 8). Consequently it was decided to investigate the matter more thoroughly.

Block Diagram of a Separately Excited D-C Motor

In order to describe completely the performance of a field control system applied to a d-c motor, it is necessary to derive the transfer functions relating the various input and output quantities associated with the machine. The drive motor is not directly involved in the basic field current regulator but with the application of current-limit control, the motor becomes an integral part of the system and the performance will depend on the properties of the motor transfer functions.

The block diagram shown in Fig. 2 represents the d-c shunt motor, with the simplifying assumption that there is no field-weakening effect due to armature reaction. The diagram is set up in terms of flux rather than field current, as this avoids any complications due to the nonlinear relationship between flux and current.

The block diagram can be described in terms of the following equations:

$$\mathcal{L}(\text{accelerating torque}) = pJ\mathcal{L}[S] \quad \ldots\ldots \quad (1)$$

$$\mathcal{L}[V-E] = R_a(1+T_{ap})\mathcal{L}[I_a] \quad \ldots\ldots \quad (2)$$

Where $\mathcal{L}[f(t)]$ is by definition the Laplace transform of the function of time $f(t)$.

The equations may be rewritten:

$$\mathcal{L}[S] = \frac{1}{Jp}\left(\frac{K_t}{\phi_f}\mathcal{L}[I_a\phi] - \mathcal{L}[M]\right) \quad \ldots\ldots \quad (3)$$

$$\mathcal{L}[V] - \frac{K_t}{\phi_f}\mathcal{L}[S\phi] = R_a(1+T_{ap})\mathcal{L}[I_a] \quad \ldots \quad (4)$$

Since these equations are nonlinear, it is not possible to write down simple transfer functions relating the input and output; and the subject of this paper has been restricted to a study of the steady-state performance. Transient effects can, however, be readily investigated by means of the computer analog described in Appendix I.

Steady-State Speed Equation

Under steady-state conditions,

$$\text{Speed, } S = \frac{\text{back-emf (electromotive force)}}{K_t\dfrac{\phi}{\phi_f}}$$

$$M/K_t\frac{\phi}{\phi_f} = \text{armature current, } I_a$$

$$R_aM/K_t\frac{\phi}{\phi_f} = IR \text{ drop}$$

$$V - R_aM/K_t\frac{\phi}{\phi_f} = \text{back-electromotive force}$$

Therefore the speed equation may be rewritten:

$$S = \frac{V - R_aM/K_t\dfrac{\phi}{\phi_f}}{K_t\dfrac{\phi}{\phi_f}} \quad \ldots\ldots \quad (5)$$

The Speed–Current Curve Under Current-Limit Control

The desirable characteristic for the current limit is a straight line between the point at which current limit commences and the value of current which is just sufficient to restore full-field excitation to the motor with the pattern input to the field current regulator for the fully weakened field condition.

Since this linear characteristic represents the maximum torque available from the machine, any current-limit curve which lies to the left of the straight line AB of Fig. 1 means that the current-limit control is restricting the motor torque below its designed maximum; and similarly, a characteristic which lies to the right of the line means that the machine is not being controlled within its design limitations.

It is therefore highly desirable that the current-limit characteristic should be a straight line, and it is equally important that the characteristic should be independent of pattern input to the field current regulator.

Calculations of Speed–Current Curve for Current Limit

Fig. 3 shows the simple system of current-limit control supplied to a motor-field current regulator. This is the system usually applied to a large reversing

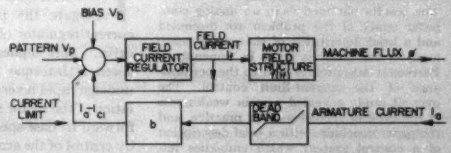


Fig. 3. Block diagram of field regulator with current-limit control

drive and it is of considerable interest to derive the mathematical equation which describes the shape of the steady-state current-limit characteristic. As interest is restricted to a steady-state analysis, no transient terms are included in the transfer functions. The current regulator itself may be assumed to have a one-to-one correspondence between the input signal and field current; this simplifies the problem without any loss of generality.

The point at which current-limit action starts will depend uniquely on the width of the deadband. For a given deadband setting, in Fig. 3, the gain b will determine the value to which the armature current will have to rise in order to restore full excitation to the motor field. It is thus apparent that the slope of the line AB in Fig. 1 is determined by the magnitude of the gain b.

This simple argument shows how the two end points of the characteristic are determined by the deadband setting and the gain in the current-limit circuit. It must be remembered that this elementary approach gives no indication of the shape of the characteristic between these two extremes, and to assume linearity would be a completely erroneous and unfounded assumption.

At this juncture it is as well to point out that with a fixed deadband setting, current-limit action is always going to start at the same point, and for a constant current-limit gain the current-limit characteristics will be parallel to one another.

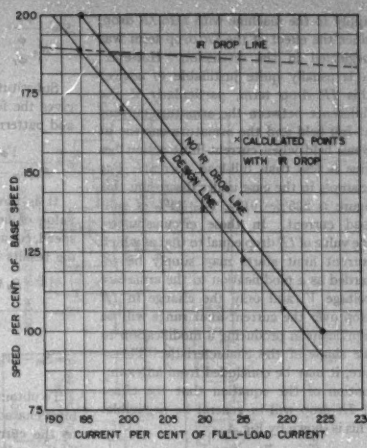
Fig. 1 shows current-limit characteristics for two intermediate speeds and it will readily be appreciated that with a fixed deadband setting the motor is being limited below its allowable maximum current. In order to ensure that current-limit action will always start at the correct point on the characteristic it is necessary to recalibrate the deadband in accordance with the pattern input to the current regulator. The incorporation of this variable deadband feature only ensures that current limit will start at the correct point; it does nothing to ensure the subsequent linearity of the characteristic once current-limit action has started.

It now remains to calculate the expression for the current-limit characteristic associated with the simple current-limited motor field regulator, as in Fig. 3.

In the block diagram, the block labelled "motor field structure" represents the functional relationship between flux and field current arising as a result of saturation of the iron. For the purpose of the calculation, machine flux ϕ will be some function, $f(I_f)$, of the field current, I_f.

The bias signal V_b ensures weak-field

Fig. 4. Curves showing the effect of IR drop on the linearity of the current-limit characteristic



full-speed excitation of the motor in the absence of any pattern signal, and V_p is the pattern signal to strengthen the motor field. Then with zero pattern signal applied to the field current regulator, the field current, for I_a greater than I_{cl}, is given by the following equation:

$$I_f = V_b + b(I_a - I_{cl})$$

and hence the machine flux ϕ will be

$$\phi = f[V_b + b(I_a - I_{cl})]$$

Having thus determined the relationship between flux and armature current, the equation to the speed–current curve is readily obtained by substituting for flux in the general speed equation 5. This gives the following result:

$$\text{Speed} = \frac{V - R_a M f\dfrac{(K_1)}{(\phi_f)}[V_b + b(I_a - I_{cl})]}{\dfrac{K_a}{\phi_f} f[V_b + b(I_a - I_{cl})]} \qquad (6)$$

It will be noted that the second term in the numerator respresents the IR drop in the armature loop for a current I_a.

In the absence of any load, the motor speed will be under the control of the pattern signal into the field current regulator. The speed will be

$$\text{Speed} = \frac{V}{\dfrac{K_a}{\phi_f} f(V_b + V_p)}$$

$$\text{Speed} = \frac{V}{\dfrac{K_a}{\phi_f} f(I_f)} \quad \ldots\ldots \qquad (7)$$

A comparison between equations 6 and 7 shows that if the IR drop term is neglected in equation 6 then the current-limit characteristic is of the same shape as the speed curve of the machine. The additional constant terms in the functional expression for field merely represent a shifting of the curve in a direction parallel to the horizontal axis.

Hence the current-limit characteristic will be curved toward the speed axis and will represent an undesirable limitation of motor torque below the maximum designed rating.

The effect of field saturation will be to increase the curvature of this characteristic. At weak field, the slope of the field-flux–field-current curve will be some six to seven times greater than the slope at full field with the iron saturated. In the case of a 2-to-1 speed-range machine, in the absence of saturation, there will be a change in slope for both the speed curve and the current-limit characteristic of 4-to-1, due to the inverse, or hyperbolic, relationship between speed and field current. In the case of the same machine with field saturation, the change in slope will be the product of the gain change due to saturation and the gain change due to the inverse relationship. For machines of the type under discussion this may well mean a slope change of the order of 30 to 1. This slope change not only indicates poor utilization of the available motor torque but it also introduces an objectionable variable gain feature into the current-limit regulator. This problem becomes increasingly significant as the speed range and degree of field saturation increase.

Strictly, the foregoing remarks only

apply to the machine with zero IR drop, since the effect of the IR drop term was neglected in the discussion. However, it is usually quite justifiable to neglect this IR drop term, particularly if the difference between the armature current at full field and the armature current at weak field when current-limit action begins is less than full-load current. For example in the case under discussion this figure is 225%–191%, i.e., 34% of full-load current. In these circumstances, the value of IR drop equal to that at which current limit starts may simply be regarded as a modification to the armature voltage V, and only the change in IR drop over the current-limit range will be significant in producing a modification to the shape of the characteristic. Therefore, it may be concluded from an examination of the equation that the effect of IR drop will in general be negligible. This is shown by Fig. 4.

Modification of the Field Current Regulator

The previous analysis shows the similarity between the current-limit characteristic and the speed curve. This suggests that if the current regulator is modified by the incorporation of some nonlinear circuit element to produce a linear relationship between speed and pattern, then a very close approximation to a linear current-limit characteristic would result. The device would have to linearize the effect due to field saturation and the effect due to the inverse relationship between flux and speed. In practice the nonlinear element may be best introduced after the pattern and current-limit signal input, before they feed the conventional current regulator. The block diagram for the linearized field regulator is shown in Fig. 5.

Graphically the nonlinear element may be described by the plot of speed against input to the linearized field current regulator and is shown in Fig. 6.

The simple equation to this curve is

$$\text{Speed } S = S_m - kV_{in}$$

where S_m is the maximum speed and k is a constant.

Under no-load conditions the machine speed S is given by:

$$S = \frac{V}{K_s \dfrac{\phi}{\phi_f}}$$

Substituting for speed in this equation gives the following relation between flux and pattern:

$$\phi = \frac{V\phi_f}{K_s(S_m - kV_{in})}$$

It is now possible to insert this new value of field flux into the equation for speed, equation 5. This gives the following results for steady-state speed:

$$S = \frac{[V - R_a M(K_s S_m - K_s k V_{in})/K_t V]}{K_s(S_m - kV_{in})} K_s V$$

$$S = S_m - kV_{in} - \frac{R_a M K_s (S_m - kV_{in})^2}{K_t V^2}$$

To obtain an expression for the current-limit characteristic, the appropriate input to the current regulator under current-limit conditions must be substituted for the input V_{in}.

This gives the following equation for speed:

$$S = S_m - kb(I_a - I_{c1}) - \frac{R_a M}{K_t \dfrac{\phi}{\phi_f}} \frac{(S_m - kV_{in})}{V}$$

$$S = S_m - kb(I_a - I_{c1}) - IR \text{ speed drop}$$

The calculated results plotted in Fig. 4 show the very small departure from linearity which results when the IR drop is taken into account.

Armature reaction will also have some effect on the shape of the current-limit characteristic. Generally this will mean that the field flux is reduced below the value of flux which would be due to field current alone. Hence a regulator system that had a linear speed characteristic under no-load conditions would be slightly nonlinear under peak load conditions. It would be possible, of course, to linearize the field-current regulator for the peak loads as dictated by the current-limit characteristic, and under these circumstances armature reaction would not effect the linearity of the characteristic. However, this effect is not likely to be significant and in practice the added complication of obtaining a linear speed characteristic under load could not be justified.

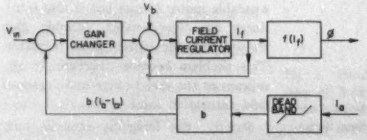


Fig. 5 (left). Block diagram of linearized field-current regulator

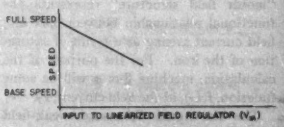


Fig. 6 (right). Linearized field-current regulator characteristic

Conclusions

In summing up it may be said that the usual technique for obtaining current-limit protection with a field current regulator does not provide the best type of current-limit characteristic. Furthermore the characteristic will be entirely dependant upon the speed range and the degree of saturation in the machine field. The large gain change in the current-limit loop presents a stability problem to the designer, and, to the user, the poor characteristic means an unnecessary limitation of the peak torque available from the drive.

The incorporation of a reasonably simple nonlinear element into the system enables the afore-mentioned difficulties to be overcome.

The advantages to be obtained in applying the linearized regulator to a drive will depend largely upon the type of service demanded from the equipment. Where the user requires to operate the drive at its peak torque rating, the system offers an important advantage. However, in cases where the output is limited either by the thermal rating of the motor or by applications which do not call for the maximum available torques in the field-weakened range, the advantages are limited to the improvements which result from the elimination of the variable gain feature from the regulator loop.

When the difficulty of loading large motors sufficiently to obtain a plot of the current-limit characteristic is appreciated, it is not surprising that the limitations in such a commonly used system should, in the past, have received so little comment. It is, however, a very simple matter to obtain a current limit characteristic from the analog computer setup described in Appendix I.

Appendix I. Description of the Analog

The following is a brief description of the analog which is shown in Fig. 7.

Amplifiers *1*, *2*, *3*, *4*, and *25* represent the motor.

The input to amplifier *2* from potentiometer *R3* represents generator voltage. Amplifier *3* is the armature time delay T_a (armature L/R) so that the output of amplifier *3* represents armature current.

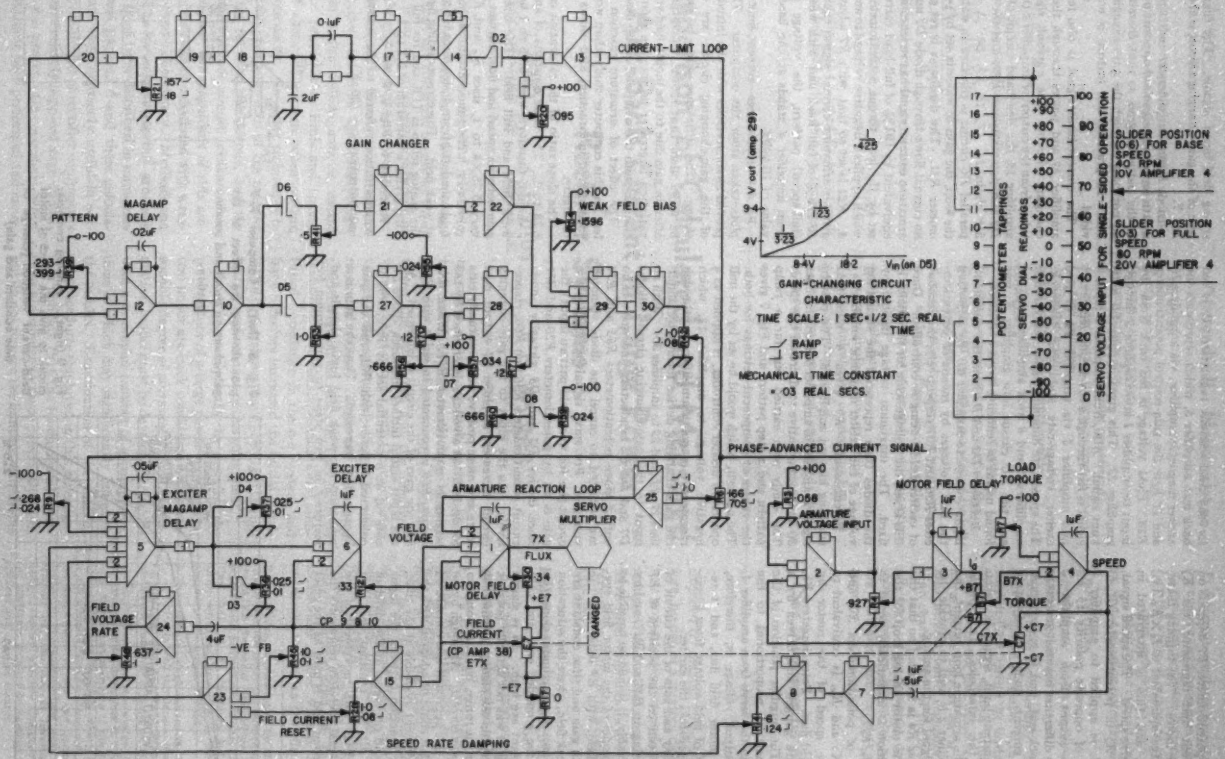


Fig. 7. Mill-motor field-current regulator analog schematic

The setting of potentiometer *B7* depends on the field current and represents the K_f constant of the motor. The tapping therefore has a voltage equivalent to $I_a \times K_f$, i.e., motor torque, and this is integrated in amplifier *4* to give speed. The speed signal, multiplied by K_s in potentiometer *C7* to give back-emf, is applied to amplifier *2* input, and is of opposite polarity to *R3* voltage. The output from amplifier *2* therefore is the difference between generator voltage and back-emf, and is proportional to armature current under steady-state conditions.

Since the current-limit signal is taken from across the interpoles, i.e., phase advanced on the armature loop current, the output of amplifier *2* is used to provide the current-limit signal for the analog.

Potentiometer *R4* is adjusted so that the gain around the motor analog loop for the full-field condition is numerically equal to $1/To$ (i.e. $K_t K_s / R_a J$).

Adjustment of potentiometers *B7* and *C7*, altering K_t and K_s, is equivalent to controlling the speed of the motor by changing its field flux. Potentiometers *B7*, *C7*, and *E7* are ganged and driven by a servomotor. The angular position of the servomotor shaft is proportional to the voltage input to the servo from amplifier *1*. Amplifier *1* output voltage is, therefore, equivalent to motor field flux, and it receives an input of field voltage from amplifier *6*, which represents the exciter.

The time delay and gain of the circuit of amplifier *1* depend on the position of *E7* slider. By setting up *R30* and *R17*, and by padding *E7* suitably, it should be possible to obtain a fair model of the motor field system. At full field, with *E7* slider near the "top" end, both gain and time constant are low; and at weak field, where the saturation curve of the motor has maximum slope, the slider is near the grounded end, and gain and time constant are high. As is shown in the following, it was not actually possible in this way to obtain the change in gain that is required over the range, but the change in time constant was correct. Although by no means perfect, this approach was considered to be the best that could be adopted with the computing equipment available at the time.

The transfer function between the input to amplifier *1* and the signal at the slider of potentiometer *E7* has constant gain but varying time delay with movement of the slider. This signal therefore represents the motor field current and is used as the reset for the field current regulator. It is fed back into amplifier *5*, which represents the exciter magnetic amplifier, via amplifiers *15* and *23* and the gain-control potentiometer *R28*. Field voltage is fed back from the input to amplifier *1* into amplifier *5* through amplifier *23*. This gives a powerful anti-hunt signal. Also, from amplifier *1* input a signal is fed through a 0.4-second *RC* circuit into amplifier *5* via amplifier *24* to give exciter voltage-rate stabilizing feedback. The speed-rate damping signal, fed through amplifiers *7* and *8* from amplifier *4* output and the *RC* circuit into amplifier *5*, was found useful in reducing speed overshoot.

Rectifiers *D3* and *D4* limit the maximum input voltage to amplifier *6*. By adjusting the bias voltages by means of potentiometers *R36* and *R37* respectively, the amount of exciter field forcing is controlled. Once this is set, the voltage scale for the analog is fixed. The maximum voltage limited by the diodes then represents the saturation voltage of the exciter magnetic amplifier.

Potentiometer *R9* gives a weak-field standing bias on amplifier *5*. The pattern signal is applied to amplifier *5* from amplifier *30* and goes negative to decrease the speed. The study was carried out for a ramp pattern and a step pattern and Fig. 7 gives the values of component settings for both. The pattern, which would be set by the mill operator's master controller, is applied as a negative voltage to amplifier *12*, which represents a magnetic amplifier, and through the gain-changing device formed by amplifiers *27*, *28*, and *29* to amplifier *30*. With no input to the gain-changing circuit, diodes *D7* and *D8* are conducting and there is no output since potentiometers *R55* and *R54* are adjusted to cancel the diode standing voltages. For small signals applied to the circuit (i.e. for weak field pattern input), the gain is set by potentiometers *R70* and *R71* since diodes *D7* and *D8* virtually short-circuit potentiometers *R56* and *R60*, respectively. As the input to the gain-changing circuit increases, amplifier *27* output goes more and more positive and eventually cuts off diode *D7* at a voltage determined by the setting of potentiometer *R57*. When this occurs, *R56* is connected in series with *R70* and the gain of the circuit is increased. As the input signal further increases, diode *D8* cuts off at a potential set by *R59*, *R60* is connected in series with *R71*, and the gain is once more increased. Thus the gain is a minimum for weak field input signals and increases in two steps to a maximum for strong field input signals. The characteristic of the gain-changing circuit is shown in Fig. 7.

Diode *D6* and amplifiers *21* and *22* provide a path for forcing signals of opposite polarity to the pattern signal, so that these signals do not pass through the gain-changing circuit. In practice this feature is not provided.

The current-limit signal is also fed into amplifier *12* and passes through the gain-changing circuit. The current signal is taken from the output of amplifier *2* through amplifier *13*, the deadband circuit of diode *D2* and potentiometer *R20*, and the amplifier chain *14*, *17*, *18*, *19*, *20*. A series/parallel *RC* circuit is effective in stabilizing the loop, the gain of which is set by potentiometer *R21*. The deadband is set so that no current flows through diode *D2* until the armature current signal has reached a value equivalent to 191% of full-load current, as set by the load torque potentiometer *R27*, which applies a negative voltage to amplifier *4*. *R21* is set so that, with *R7* adjusted to give an armature current signal equivalent to 225% of full load, and with weak-field (full-speed) pattern, the voltage output from amplifier *4* stands at the base-speed value of 10 volts.

Fig. 8 reproduces current-limit *X-Y* plots taken from the computer. The improved characteristics with the introduction of the gain-changing circuit are clearly shown. The difference will, in practice, be wider than that shown, since the variation in gain due to the motor field-strength variation is greater in practice than is obtainable with the computer setup described.

The effect of armature reaction (again less on this computer study than would be encountered in practice) also shows up in the figure.

The slope ratio of the saturation curve is 5.75 to 1. The time constant of amplifier *1* should, therefore, vary 5.75 to 1 between weak field and full field. The change in gain between speed and field current should vary 5.75×4 to 1, since K_s and K_t both vary 2 to 1. The time constant of the motor field calculated from design data is 6 seconds at weak field. The corresponding figure for the analog setup is 11.6, and for full field is 2.07, giving a variation of 5.7 to 1.

Referring to Fig. 9, k, m, n, q, r are resistance ratios, referred to the resistance of one potentiometer (30 *K*).

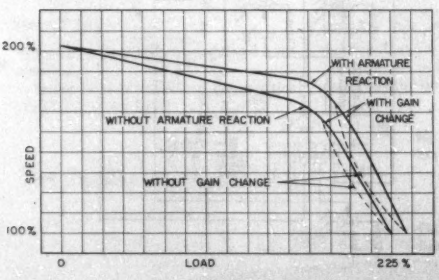


Fig. 8 (left). Computer X-Y plot showing the effect of the gain-changing circuit and armature reaction

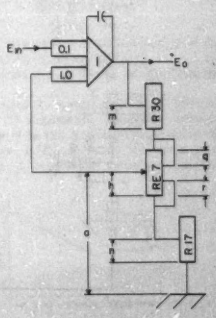


Fig. 9 (right). Analog computer circuit to give the effect of motor field-current variation (change of time constant and gain)

$$a=\frac{h+m+n}{m+(1-q-r)+n}$$

If s is the output of amplifier 4, then $h=c/s$ where c is a constant and equal to 6 in this case. Also

$$k=\frac{E_g}{100}$$

and

$$(10E_{in}-aE_o)\frac{1}{p}=E_o$$

From these equations it follows that:

$$E_{in}=\frac{\frac{10s^2}{s^2}+\frac{(n-r)10c}{s}}{m+(1-q-r)+n}$$

and differentiating with respect to s,

$$\frac{dE_{in}}{ds}=\frac{10c}{m+(1-q-r)+n}\frac{1}{s^2}\left[2c+(n-r)s\right]$$

and the ratio of the gains of E_{in} to s at weak field and full field is 10.3 to 1 for the setup used.

Appendix II. Nomenclature

VARIABLES

E = the motor back-emf
I_a = armature current
ϕ = motor field flux
S = motor speed (radians per second)
s = analog voltage corresponding to speed
I_f = field current
M = load torque
a, h = analog potentiometer ratios

CONSTANTS

V = voltage applied to the armature
K_1 = torque per ampere for full-field flux
K_2 = back-emf per radian per second for full-field flux
R_a = resistance of motor armature, compensating windings, and interpoles
J = inertia of armature and load (slug square feet)
ϕ_f = full-field flux
T_a = armature loop time constant
T_s = mechanical time constant at full field
p = Laplace transform p
S_m = maximum no-load (weak-field) speed
c, k = constants

V_b = current-regulator bias signal voltage
b = current-limit loop-gain constant
V_p = current regulator pattern signal voltage
I_{d1} = the value of armature current which is just sufficient to overcome the deadband in the current-limit loop
m, n, q, r = analog potentiometer ratios

DEFINITIONS

Speed curve: A plot of no-load motor speed against motor field current.

Current-limit characteristic: If a motor with a field-current regulator with an overriding current-limit circuit is loaded and a plot made of speed against armature current, that part of the plot which is determined by an armature current signal being injected into the field control regulator is the current-limit characteristic of the system.

IR drop: The potential drop across the motor armature, interpoles, and compensating windings due to the armature loop current.

(See Discussion on page 692)

Comparison of Calculated and Measured Arc-Back Current in Large Power Rectifier Systems

J. TENO
ASSOCIATE MEMBER AIEE

C. H. TITUS
MEMBER AIEE

R. N. WAGNER
MEMBER AIEE

EVEN with the high degree of reliability which mercury-arc rectifiers possess today, there are still a few occasions when current is carried in the wrong direction in one of the rectifier tanks and hence an arc back is created. Since this arc-back current can rise to an abnormally high value in a few milliseconds, a means must be provided to limit and to cut off this abnormal current when an arc back does occur. In order to be able to specify the characteristics of the rectifier tanks, the rectifier transformer, and the anode and cathode circuit breakers, it is essential that the magnitude of this arc-back current can be calculated with a reasonable degree of accuracy. Methods and procedures for calculating this arc-back current have been described in numerous books and papers,[1-6] and these calculations have been extremely useful in permitting the design of rectifier equipments which have reliably supplied large quantities of d-c power. On occasions when field tests have been performed and measurements of this arc-back current have been made, an abnormally large discrepancy has usually been found to exist between the calculated and the measured values of this current with the calculated values being appreciably higher than the measured values. While the calculated results were on the safe side, they have, on occasions, required the installation of additional impedance in the a-c system and have therefore increased the initial installation cost and may have penalized the operating performance of this equipment.

In order analytically to describe the performance of a large power rectifier system under arc-back conditions with a reasonable degree of accuracy, a good mathematical model must be developed, coupled with a minimum number of assumptions, none of which may significantly influence the results. Examination of the analytical procedures employed in this investigation indicate that the fundamental equations are similar to those previously used, but do not require the limiting restrictions imposed by manual calculations. Therefore, by use of a modern digital computer it is now practical to perform the kind of analyses and calculations which would have been impractical if they were performed manually with a conventional desk calculator.

Nature of the Investigation

The nature and purpose of this investigation can be described as being fourfold:

1. To describe accurately, by analytical means, the magnitude and waveshape of the abnormal transient currents which occur in both the low-voltage a-c circuits of a large double-Y power rectifier, as well as the transient direct current which is supplied by the external system when a power rectifier tank arcs back. This analysis must take into account the nonlinear characteristics of the interphase transformer under various conditions of unbalance as well as the effect of arc drop

Paper 58-223, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 2-7, 1958. Manuscript submitted November 6, 1957; made available for printing June 5, 1958.

J. TENO is with Lehigh University, Bethlehem, Pa.; C. H. TITUS is with the General Electric Company, Philadelphia, Pa.; and R. N. WAGNER is with the Aluminum Company of America, Pittsburgh, Pa.

The authors wish to acknowledge the contributions of their principal associates in this investigation: L. J. Harris and members of the Aluminum Company of America Rockdale plant; A. Schmidt, Jr., of the General Electric Company, who assisted in the preparation of the test program and made numerous recommendations and suggestions for carrying out the over-all investigation; W. A. Abbott, General Electric Company, who provided the assistance required to conduct the tests in an efficient and effective manner.

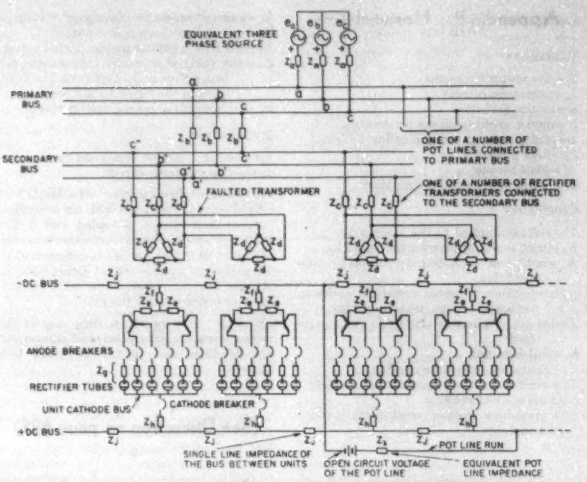

EQUIVALENT THREE PHASE SOURCE
PRIMARY BUS
SECONDARY BUS
ONE OF A NUMBER OF POT LINES CONNECTED TO PRIMARY BUS
ONE OF A NUMBER OF RECTIFIER TRANSFORMERS CONNECTED TO THE SECONDARY BUS
FAULTED TRANSFORMER
-DC BUS
ANODE BREAKERS
RECTIFIER TUBES
UNIT CATHODE BUS
CATHODE BREAKER
+DC BUS
POT LINE RUN
SINGLE LINE IMPEDANCE OF THE BUS BETWEEN UNITS
OPEN CIRCUIT VOLTAGE OF THE POT LINE
EQUIVALENT POT LINE IMPEDANCE

Fig. 1. Aluminum reduction rectifier system used for investigation

and impedance introduced by the protective equipment while interrupting the arc-back current.

2. To measure the transient impedance characteristics of the more important circuit elements of a large double-Y mercury-arc rectifier system in order to determine the correct system constants to be used in the analytical phase of this investigation.

3. To initiate arc backs under carefully controlled conditions with equipment which would present substantially the same nonlinear impedance characteristics as that imposed by natural arc backs under the most severe conditions encountered in service.

4. To measure the available current which would be encountered if high-speed protective equipment were not available and then to show the effect of high-speed anode breakers on limiting both the magnitude and duration of the arc-back current.

Description of the Power System Under Investigation

Through the assistance and co-operation of the Aluminum Company of America, it was possible to make system measurements and arc-back tests on a large d-c electrochemical system of the type used for feeding the pot lines of an aluminum reduction plant. (Fig. 1 shows the equivalent power system which was studied.) While there are several pot lines supplied from the same a-c bus, attention will be focussed on the performance of one pot line which consists of 12 6,000-ampere double-Y pumped-type mercury-arc rectifiers capable of supplying 750 volts to the d-c bus. The rectifier equipment under investigation is located at the extreme end of the d-c bus but by closing the d-c bus tie equipment, it was possible to tie a second pot line of similar rating to the pot line under investigation and thereby have a system in which the rectifier undergoing arc-back test could be located in the middle of a d-c system having twice the capacity of a single pot line.

In addition, each rectifier transformer used to supply two 6,000-ampere rectifiers was equipped with current-limiting reactors, one in series with each of the 13.8-kv primary transformer phases. The tests were so arranged that these reactors could either remain in series or be removed from the system, thereby modifying the magnitude of the arc-back current by controlling the primary impedance of the rectifier equipment under test. Provision was also made for the interphase transformer of the rectifier under test to be saturated in either one direction or the other before the arc back was initiated. This equipment also permitted all tanks to fire normally during the arc back in order to obtain an arc back of maximum severity under each condition of interphase saturation.

Description of the Analytical Work

The complete résumé of the mathematical analysis and techniques used for solving this complex electric system on a digital computer have been described.[7] In order to make this analysis as rigorous as possible, considerable effort was expended in reducing the assumptions to the conditions usually encountered in practice or to those having only a negligible influence on the comparative results.

The following assumptions were made:

1. The excitation of the a-c generator remained constant during the arc-back period.

2. All impedances in the system were linear except the magnetization characteristic of the interphase transformer and the arc drop of the anode breaker.

3. All arc backs were initiated as full commutation arc backs, thereby permitting maximum current to flow.

4. The rectifier under test was operated at full phase rather than being used with phase-retarded firing.

5. The transformer, although used to feed two rectifiers, has such a low coupling factor between the sections feeding each rectifier that it can be considered as two separate transformers housed in one tank.

6. In the case where a common 1-ohm reactor was connected in series with the primary of the rectifier transformer, it was assumed that the voltage would dip so low due to the drop in this common reactance that the other rectifier on the same transformer was not providing any d-c contribution to the bus and that the d-c bus had the equivalent of 10 rather than 11 rectifiers feeding back into the fault.

7. It was assumed that the magnetization curve of the magnetic material of the interphase transformer could be represented by three linear approximations. While it is recognized that this is not completely accurate, the technique employed provided very satisfactory results.

8. It was assumed that the pot-line current would remain constant during the arc back.

The mathematical analysis describes the current in the phase having the arc back, the current in the contributing phases of the same Y, the current contributed by the adjacent Y of the same transformer, and the current supplied from the d-c bus to the phase having the arc back. Verification of the accuracy of these calculated results is shown by the good agreement obtained between the measured and calculated results.

The type of analysis and the general techniques for programming equations of the type employed are applicable to several types of digital computers, but the computer employed in this investigation was an International Business Machines Corporation digital type 650 of a style frequently employed in payroll and accounting activities.

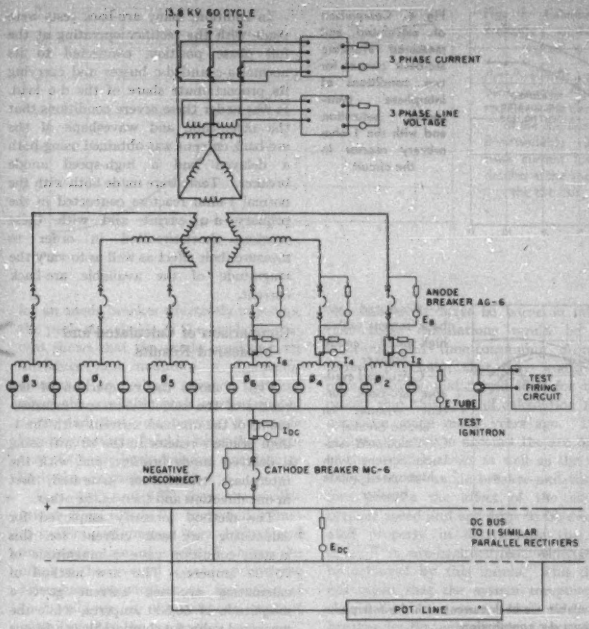


Fig. 2. Test circuit using a 6,000-ampere 750-volt rectifier

Description of the Test Program

While planning the test program, it became evident that the arc backs should be synchronously controlled so that each test would provide the highest available current that could be obtained under natural prevailing conditions. This was accomplished by connecting a 16-inch pumpless rectifier tank in the inverse direction across one of the tanks of the rectifier equipment under test, and then firing this tank with a single-shot phase-controlled tube-type firing circuit. By using a rectifier tank to initiate the arc back, the arc drop of this tank is probably comparable to that obtained under natural arc-back conditions. While it is recognized that this may not always be exactly correct, the difference in arc drop would probably have a negligible effect on both the waveshape and the magnitude of the arc-back current.

It also became evident that the relative saturation in the interphase transformers could have a considerable influence on the instantaneous amount of direct current which could be fed back from the d-c system through the arc-back path. If this interphase transformer is saturated in one direction it may produce a considerable delay until the potential of the d-c system can reverse the magnetization of the core and drive it into saturation in the reverse direction. If the interphase transformer were saturated in the other direction at the inception of the arc back there would be no significant time delay and consequently the instantaneous magnitude of the direct current would reach a higher value. In order to control the direction of saturation of the interphase transformer, excitation of the igniter circuit was removed from the rectifier tanks of one phase in either one Y or the other by means of synchronously controlled thyratrons and these rectifier tanks were rendered inoperative until just prior to the inception of the arc back at which time all tanks were permitted to fire, thereby providing maximum available arc-back current.

Fig. 2 shows the power, measuring, and control equipment used to initiate the arc back, to control interphase transformer saturation, and to measure the various transient phenomena taking place during the arc-back tests.

Inasmuch as the arc-back current is calculated under the most severe conditions without the current limiting provided by high-speed anode breakers, one pole of the anode breaker was modified so that it would not part its contacts until after the initial peak of arc-back current was attained. This then would permit a direct comparison between the measured and calculated values of arc-back current.

A more detailed description of the test program, with particular emphasis on the segment covering the anode breaker tests, is given in a paper by Harris, Scully, and Stewart.[5]

In order to measure the magnitude of the equivalent impedance in the a-c system through the arc-back path without d-c contribution, an arc back was initiated by the short-circuiting rectifier tank with the d-c cathode breaker open prior to the test. This permitted the

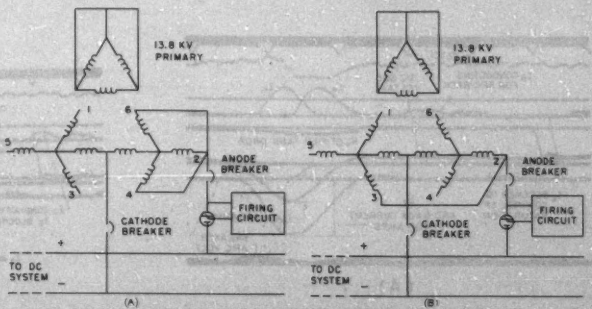


Fig. 3. Circuits used to determine rectifier equipment constants

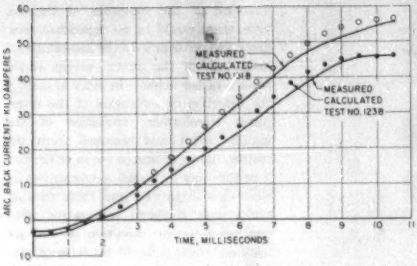
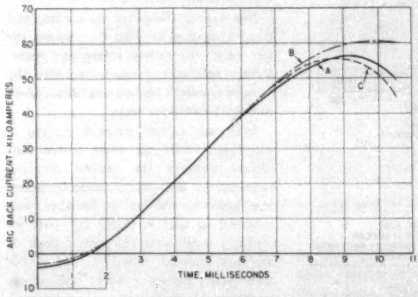


Fig. 4. Comparison of calculated and measured available arc-back current for two conditions of interphase transformer saturation and with the 1-ohm primary reactor in the circuit



A—Measured arc-back current with a high-speed anode breaker
B—Calculated available arc-back current without an anode breaker
C—Calculated arc-back current with a high-speed anode breaker

Fig. 5. Comparison of calculated and measured available arc-back current with the interphase transformer saturated for maximum d-c contribution

magnitude and waveshape characteristics of the a-c component of current to be measured without the effect of d-c feedback.

By disconnecting the primary of the rectifier transformer from the 13.8-kv bus, short-circuiting the three primary phases together, and then by connecting the secondary phases together as indicated in Fig. 3(A) and 3(B), it was possible to obtain the effect of the d-c contribution independent of the a-c contribution. In these cases the power was supplied from the d-c bus and the fault initiated by the mercury-arc short-circuiting tank. While the results of these tests do not lend themselves to direct superposition, they permit the equivalent impedances of the various system elements to be determined by a combination of measurement and analysis.

In addition, many arc-back tests were made with the rectifier operating at the full phase position, connected to its normal a-c and d-c busses and carrying its proportionate share of the d-c load. It was under these severe conditions that the magnitude and waveshape of the arc-back current was obtained using both a delayed and a high-speed anode breaker. Tests were made both with the normal 1-ohm reactors connected in the primary a-c circuit and with these reactors short-circuited in order to measure their effect as well as to vary the magnitude of the available arc-back current.

Comparison of Calculated and Measured Results

Fig. 4 shows the comparison of the results of the calculations and measurements of the arc-back current with the 1-ohm primary reactor in the circuit, using a delayed anode breaker, and with the interphase transformer saturated, first in one direction and then in the other.

The method formerly employed for calculating arc-back current for this system condition gave a magnitude of 75,700 amperes. The new method of calculating arc-back current gave a magnitude of 56,500 amperes, while the measured value for comparable conditions was 55,000 amperes.

Fig. 5 shows the calculated value of the available arc-back current with the interphase transformer saturated in the direction to give the maximum amount of d-c contribution, and with 1 ohm removed from the primary circuit of the rectifier transformer. This figure also shows a comparison between the measured value and the calculated value of the arc-back current when a high-speed anode breaker is used.

Investigation of the criteria necessary

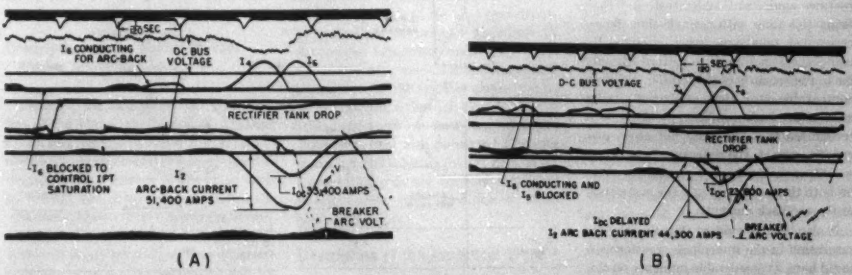


Fig. 6. Typical oscillograms taken during arc-back tests with the interphase transformer saturated to give (A) maximum and (B) minimum d-c contribution

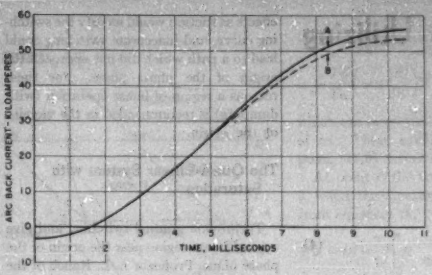


Fig. 7. Calculated available arc-back current

A—Available arc-back current for a rectifier at the center of the d-c bus

B—Available arc-back current for a rectifier at the end of the d-c bus

for an anode breaker effectively to reduce the peak magnitude of the arc-back current shows that the anode breaker must introduce and maintain a significant amount of arc voltage prior to the time when the available peak current would occur and that just parting the anode breaker contacts quickly after the inception of the arc back does not significantly limit the arc-back current.

Figs. 6(A) and 6(B) show the oscillograms from which the measured values of system performance were derived and an examination of curve *F* shows the effect of the reversal of saturation of the interphase transformer on the delay of the build-up of d-c cathode current contributed to the arc back.

Fig. 7 shows the comparison of the calculated arc-back current for a rectifier at the end of the d-c bus in comparison with the calculated current in the arc back for a comparable rectifier located at the middle of the d-c bus.

Conclusions

Since the calculated and measured results are in good agreement, it is apparent that the mathematical model is correct and that the assumptions are reasonable. It must be borne in mind that these calculations would be so laborious and time-consuming if performed manually, that they would be impractical to make, and it is for this reason that this type of calculation had not been made many years ago. This analysis takes into account the arc drop of the rectifier tanks as well as the arc voltage of the anode breaker and therefore permits the effect of the anode breaker speed and arc drop to be evaluated properly in determining the reduction in arc-back current which can be achieved by this means. This does not mean that the system components should be chosen on the basis of anode breaker let-through current rather than system available current without the anode breaker functioning. Examination of this investigation makes it apparent that the speed of parting of anode breaker contacts alone is not the answer and that the current is not limited until an appreciable magnitude of arc voltage is maintained across the breaker contacts. The analysis of measurements also indicates the importance of the relative magnitude and direction of flux in the core of the interphase transformer. This is probably one of the components which has accounted for some of the variations that have been observed in arc-back and cathode current during previous tests.

Since this mathematical analysis has exhibited an acceptable degree of accuracy when compared with measurements it will be extremely useful in calculating the relative magnitude of the arc-back current as well as the forward currents contributed by the other phases of the rectifier during this transient condition, and thereby can be used in determining the proper impedances and the system configurations required for new or modified power rectifier systems.

It is apparent that this over-all program was extremely ambitious both from an analytical and a measurements point of view, and probably entailed one of the most comprehensive series of tests that has ever been made on a large mercury-arc rectifier system. It is therefore believed that this investigation will aid the technical industry committees in their work.

References

1. Analysis of Rectifier Circuits. E. F. Christensen, C. H. Willis, C. C. Herskind. *AIEE Transactions*, vol. 63, 1944, pp. 1048–58.

2. Rectifier Fault Currents. C. C. Herskind, H. L. Kellogg. *Ibid.*, vol. 64, Mar. 1945, pp. 145–50.

3. Mercury Arc Power Rectifiers (book). O. K. Marti, H. Winograd. McGraw-Hill Book Company, Inc., New York, N. Y., 1927.

4. Rectifier Fault Currents—II. C. C. Herskind, A. Schmidt, Jr., C. E. Retzig. *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 243–53.

5. Arc-Backs in Rectifier Circuits: Artificial Arc-Back Tests. R. D. Evans, A. L. Maslin. *Ibid.*, vol. 64, June 1945, pp. 303–11.

6. Rectifier Arc-Back Study on the Analogous Computer. J. K. Dillard, C. J. Baldwin, Jr. *Ibid.*, vol. 73, pt. I, July 1954, pp. 198–208.

7. A Mathematical Model and Procedure for Arc-Back Current Calculations for Power Rectifiers. H. P. Fullerton, J. Teno. *Ibid.*, vol. 77, pt. II, Nov. 1958, pp. 456–64.

8. Anode Breaker Field Testing with a High Capacity Rectifier System. L. J. Harris, T. J. Scully, V. N. Stewart. *Ibid.*, vol. 78, pt. II, 1959 (*Paper no. 58-224*).

A Dual-Mode Servomechanism Utilizing Saturation Switching

H. R. WEED
ASSOCIATE MEMBER AIEE

F. C. WEIMER
MEMBER AIEE

Synopsis: This paper discusses the theoretical analysis and experimental verification of a dual-mode quasi-linear servomechanism. After a brief review of the operation of the simple dual-mode servo system, the desirability of linear operation for small errors is pointed out. A method of producing a gradual transition from one mode of operation to the other and finally to nearly linear operation is investigated.

The theoretical analysis makes use of the phase plane to show the three distinct regions of operation and then the construction of a complete trajectory is discussed and plotted.

The description and test results of an experimental system incorporating a specialized computer to accomplish the transition are included to demonstrate the improvement in time response of the system as the quasi-linear region is made small compared to the input signal.

THE CONVENTIONAL dual-mode servo system has been described in a paper by K. C. Mathews and R. C. Boe.[1] The output member is assumed to have a moment of inertia J, negligible friction, and is driven by a motor with maximum torque T_m. In the dual-mode system the torque of the motor can be only $+T_m$, 0, or $-T_m$.

Thus,

$$J\ddot{C} = \pm T_m \tag{1}$$

where C is the angular position of the output member and each dot indicates differentiation with respect to time. If the acceleration of the input R is zero (as in the case of a step or ramp function), the error acceleration is

$$\ddot{E} = \ddot{R} - \ddot{C} = -\ddot{C} \tag{2}$$

Thus equation 1 becomes

$$\ddot{E} = \frac{d\dot{E}}{dt} = \mp \frac{T_m}{J} \tag{3}$$

It is also true that

$$\frac{dE}{dt} = \dot{E} \tag{4}$$

Dividing equation 3 by equation 4 gives

$$\frac{d\dot{E}}{dE} = \mp \frac{T_m}{J\dot{E}} \tag{5}$$

or

$$\dot{E}d\dot{E} = \mp \frac{T_m}{J} dE \tag{6}$$

Equation 6 may be integrated to give

$$\dot{E}^2 \pm \frac{2T_m}{J} E = K \tag{7}$$

where K is a constant of integration determined by initial conditions. The families of parabolas for $\pm T_m$ represented by equation 7 are shown in the phase plane, Fig. 1. The two parabolas represented by

$$\dot{E}^2 \pm 2\frac{T_m}{J} E = 0 \tag{8}$$

pass through the origin of the phase plane. The branches of these curves in the second and fourth quadrants are directed toward the origin and would cause the system to come to rest with zero error. These two branches together form the switching curve which is shown by heavy lines in Fig. 1. The equation of the switching curve is

$$\dot{E}|\dot{E}| + 2\frac{T_m}{J} E = 0 \tag{9}$$

For ideal operation the system should follow the parabola that passes through the initial condition point until it intersects the switching curve. Then the torque should reverse and the system should follow the switching curve to rest with zero final error. A typical path is shown by the dotted line in Fig. 2. In this case there is an initial error as a result of a step input. Examination of the path reveals that full positive torque should be applied until the error is reduced to half its initial value, and then full negative torque should be applied until the system comes to rest with zero final error.

The switching from positive to negative torque could be accomplished by means of a specialized computer. Such effects as friction would modify the switching curve, and inaccurate switching could lead to a path which did not approach the origin of the phase plane. For these reasons a region of linear operation (with damping) is recommended in the vicinity of the origin.

The Quasi-Linear System with Saturation

To provide gradual switching and give a quasi-linear region near the origin of the phase plane, Professor L. L. Rauch of the University of Michigan suggested that the motor torque should obey the function suggested by equation 10. This equation states that

$$T = T_m \operatorname{sat}\left\{\frac{1}{W}\left[E + \frac{J}{2T_m}\dot{E}|\dot{E}| + W \operatorname{sat}\frac{a\dot{E}}{W}\right]\right\} \tag{10}$$

where the saturation function is defined as follows:

$$\operatorname{sat} x = \begin{cases} -1 \text{ for } x \leq -1 \\ x \text{ for } -1 \leq x \leq 1 \\ 1 \text{ for } x \geq 1 \end{cases} \tag{11}$$

The design parameters W and a will be discussed later.

Because of the two saturation functions in equation 10 there will be three distinct types of operation as indicated by the three regions of the phase plane shown in Fig. 2. In region 1 the argument of equation 10 is greater than 1, as shown by

$$\left|\frac{1}{W}\left[E + \frac{J}{2T_m}\dot{E}|\dot{E}| + W \operatorname{sat}\frac{a\dot{E}}{W}\right]\right| > 1 \tag{12}$$

so that

$$T = \pm T_m \tag{13}$$

and the operation will be the same as for the true dual-mode system. In region 2 only the error-rate term is saturated so

$$\left|\frac{a\dot{E}}{W}\right| > 1 \tag{14}$$

and

$$T = \frac{T_m}{W}\left[E + \frac{J}{2T_m}\dot{E}|\dot{E}| \pm W\right] \tag{15}$$

In region 3 neither term is saturated so

$$T = \frac{T_m}{W}\left[E + \frac{J}{2T_m}\dot{E}|\dot{E}| + a\dot{E}\right] \tag{16}$$

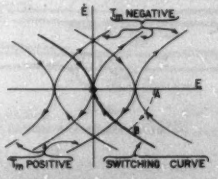


Fig. 1. Phase plane for dual-mode servo

Paper 58-1191, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted October 28, 1957; made available for printing, August 5, 1958.

H. R. Weed and F. C. Weimer are with Ohio State University, Columbus, Ohio.

The authors gratefully acknowledge the financial support of this project by the Graduate School of Ohio State University and the work of R. P. Halverson, F. G. Kilmer, and G. C. Milligan who contributed to the project while completing their theses for the Master of Science degree.

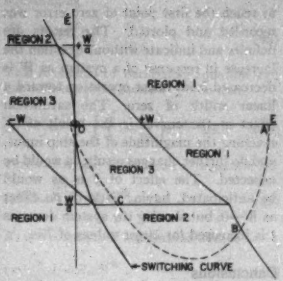


Fig. 2. Regions of the phase plane

If $\dot{E}$ is sufficiently small equation 16 becomes

$$T=\frac{T_m}{W}[E+a\dot{E}] \tag{17}$$

and the operation approaches linearity. These three regions with their boundaries and the previously defined switching curve are shown in Fig. 2. It can be seen that W is half the width of quasi-linear region, and a is the coefficient of the linear error-rate damping term.

Phase-Plane Trajectories

It is possible to calculate the trajectory of a system following equation 10. A typical path originating in region 1, is shown in Fig. 2. The trajectory from point A to point B lies in region 1 and follows the parabolic path of equation 7. At point B the traejctory enters region 2 where the equation of motion is derived from equation 15.

$$T=-J\ddot{E}=\frac{T_m}{W}\left[E+\frac{J}{2T_m}\dot{E}|\dot{E}|-W\right] \tag{18}$$

Equation 18 is reduced to a form of Bernoulli's equation and integrated to give

$$\dot{E}^2=\frac{2T_m}{J}E+K_1e^{E/W} \tag{19}$$

where K_1 is a constant determined by the values of E and $\dot{E}$ at point B. For plotting it may be convenient to use the method of isoclines. Since

$$\ddot{E}=\frac{d\dot{E}}{dt}=\frac{d\dot{E}}{dE}\frac{dE}{dt}=\dot{E}\frac{d\dot{E}}{dE} \tag{20}$$

Equation 18 becomes

$$-J\dot{E}\frac{d\dot{E}}{dE}=\frac{T_m}{W}E+\frac{J}{2W}\dot{E}|\dot{E}|-T_m \tag{21}$$

where $d\dot{E}/dE$ is the slope of the trajectory in the phase plane and will be assigned a value m for any given isocline. Thus the equation of an isocline is

$$-J\dot{E}m-\frac{J}{2W}\dot{E}|\dot{E}|=\frac{T_m}{W}(E-W)$$

$$\text{or } -\frac{JW}{T_m}\left(m+\frac{|\dot{E}|}{2W}\right)\dot{E}=E-W \tag{22}$$

which can be plotted readily for any value of m. These isoclines go through the point $(W,0)$ rather than the origin.

At point C the trajectory enters region 3, where the equation of motion is derived from equation 16. Thus,

$$T=-J\ddot{E}=\frac{T_m}{W}\left[E+\frac{J}{2T_m}\dot{E}|\dot{E}|+a\dot{E}\right] \tag{23}$$

Equation 23 may be integrated only in special cases such as $a=0$. If $a=0$, equation 23 becomes

$$-J\ddot{E}=\frac{T_m}{W}\left[E+\frac{J}{2T_m}\dot{E}|\dot{E}|\right] \tag{24}$$

Equation 24 is the same as equation 18 except that E is replaced by $E+W$. Thus

$$\dot{E}^2=\frac{2T_m}{J}(E+W)+K_2e^{E/W} \tag{25}$$

In general the method of isoclines must be used to solve equation 23. Substituting equation 20 into equation 23 gives

$$-J\dot{E}\frac{d\dot{E}}{dE}=\frac{T_m}{W}E+\frac{J}{2W}\dot{E}|\dot{E}|+a\dot{E} \tag{26}$$

For a given isocline the slope $d\dot{E}/dE=m$.

$$\text{Thus } -J\dot{E}m-\frac{J}{2W}\dot{E}|\dot{E}|-a\dot{E}=\frac{T_m}{W}E$$

$$\text{or } -\frac{JW}{T_m}\left(m+\frac{a}{J}-\frac{|\dot{E}|}{2W}\right)\dot{E}=E \tag{27}$$

These isoclines go through the origin and approach straight lines as $\dot{E}$ goes to zero.

For small values of $\dot{E}$ equation 17 may be used, giving

$$-J\ddot{E}=\frac{T_m}{W}[E+a\dot{E}] \tag{28}$$

$$\text{or } \ddot{E}+\frac{T_ma}{JW}\dot{E}+\frac{T_m}{JW}E=0 \tag{29}$$

This represents a linear system with undamped natural angular frequency

$$\omega_n=\sqrt{\frac{T_m}{JW}} \tag{30}$$

and damping ratio

$$\zeta=\frac{a}{2}\sqrt{\frac{T_m}{JW}} \tag{31}$$

It must be noted, however, that because of the term containing $|\dot{E}|$ in equation 27, the actual damping is always more than would be indicated by equation 31.

Under appropriate conditions a trajectory may originate in region 3 or go directly from region 1 to region 3 without passing through region 2.

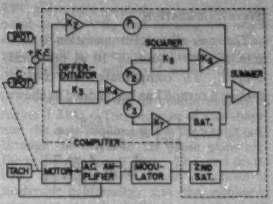


Fig. 3. Block diagram of dual-mode system

Effect of Varying Parameters W and a (or ζ)

It can be seen from Fig. 1 that W is half the width of the quasi-linear regions 2 and 3. Making W smaller would cause the system to approach the true dual-mode operation described in the first part of this paper. Increasing W would cause more of the trajectory to be in the quasi-linear region. If W is made larger than the magnitude of a given step-function input, the trajectory lies entirely within quasi-linear region 3.

Varying a merely controls the linear damping ratio ζ as shown by equation 31. It has already been pointed out that the total damping is always greater because of the term containing $|\dot{E}|$ in equation 27. This nonlinear damping is sufficient to make the system stable even if $a=\zeta=0$.

Experimental System

An experimental system incorporating the quasi-linear region was constructed. The block diagram of the system is shown in Fig. 3. Here the system is divided into the error detecting system, the computer, the output modulator, and the amplifier and compensated motor. To aid in future calculations each block indicates the symbol of its fixed gain or adjustable potentiometer setting.

For simplicity of operation the error detector consists of two 10-turn potentiometers excited from a source of d-c voltage. The d-c error signal is then fed into the computer where the operations of amplification, differentiation, squaring, and saturation are performed.

The output of the computer is then fed into a synchronized chopper modulator before power amplification. The power amplifier is a conventional a-c amplifier with multiple inputs and separate gain adjustments. Since the analytical theory proposed assumed a frictionless motor, it was desirable to compensate for a major portion of the motor's viscous-friction characteristic by use of positive tachometer feedback.

System Operation

The over-all system operation is characterized by equation 10 as previously defined. To compensate for coulomb friction a term F has been added as shown:

$$T = T_m \text{ sat} \times \left\{ \frac{1}{W} \left[E + \frac{J\dot{E}|\dot{E}|}{2(T_m + F)} + W \text{ sat} \left(\frac{a\dot{E}}{W} \right) \right] \right\} \quad (32)$$

For purposes of system adjustment, it is desirable to observe the input signal to the summer which is also the signal upon which the second saturator and hence the output amplifier operates. The input to the summer is shown in equation 33.

Summer Input

$$= \frac{1}{W} \left[E + \frac{J\dot{E}|\dot{E}|}{2(T_m + F)} + W \text{ sat} \left(\frac{a\dot{E}}{W} \right) \right] \quad (33)$$

It is also possible to express the input to the summer in terms of the symbols of Fig. 3.

Thus,

$$\text{Summer Input} = K_1K_2P_1E + K_1K_2K_4K_7P_3\dot{E} + (K_1K_2K_4P_2)^2K_3K_4\dot{E}|\dot{E}| \quad (34)$$

If like terms of equation 33 and equation 34 are compared, three necessary relationships result. They are

$$K_1K_2P_1 = \frac{1}{W} \quad (35)$$

$$K_1K_2K_4K_7P_3 = \frac{a}{W} \quad (36)$$

$$\text{and } (K_1K_2K_4P_2)^2K_3K_4 = \frac{J}{2W(T_m + F)} \quad (37)$$

From equations 35, 36, and 37 it is possible to solve for the required potentiometer settings in terms of W, a, J/T_m, F and the various K's of the system.

$$\text{Thus } P_1 = \frac{1}{K_1K_2W} \quad (38)$$

$$\text{and } P_2 = \frac{\left[\frac{J}{2K_3K_4W(T_m + F)} \right]^{1/2}}{K_1K_2K_4} \quad (39)$$

$$\text{and } P_3 = \frac{a}{K_1K_2K_4K_7W} \quad (40)$$

The independent variables of the system are to be W and ζ, where W is the half width of the linear region and ζ is the system-damping ratio in the linear region of operation near rest. The term W enters equations 38 through 40 directly while the damping ratio ζ enters through its relation to a as defined by

$$\zeta = \frac{a}{2} \left[\frac{T_m + F}{JW} \right]^{1/2} \quad (41)$$

For the system under test the fixed values were determined to be:

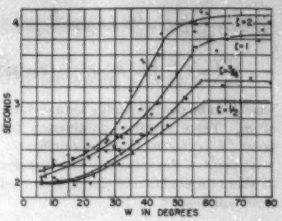


Fig. 4. Time to reach zero error versus W

$K_2 = 12.6$
$K_3 = 0.0082$
$K_4 = 18.5$
$K_5 = 0.022$
$K_6 = 78$
$K_7 = 52$
$T_m = 0.4$ inch-ounce
$F = 0.1$ $T_m = 0.04$ inch-ounce
$T_m/J = 6.2$ radians per second squared
$K_1 = V$ per 20π volts per radian
$V =$ voltage applied to d-c potentiometers

In general, the potentiometer settings should fall between 1.0 and 0.1 and this was made possible by certain adjustment of terms such as V.

Test Procedure

A step input of one radian (57.3 degrees) was chosen since this permitted the system to be tested for values of W ranging from ten per cent of the input to values greater than the input.

The test procedure was established so that it included the following experimental data.

1. A family of curves showing error versus time for a one-radian step input with ζ fixed and W varied in steps from W equal to ten per cent of the input to W larger than the input step. These data were recorded on a Sanborn model *51-600 T* recorder.

2. A family of curves showing error versus time for a one-radian step input with W fixed and ζ varied over a range from $\zeta = 0.5$ to $\zeta = 2.0$.

3. From the recorded data of 1 and 2 it was possible to plot a family of curves indicating the time necessary for the system to first reach zero error as a function of ζ and W. In an attempt to eliminate the slight variations in operation from trial to trial, each point on such a family of curves would be the average result of several successive trials.

Experimental Results

The results of the experimental data are shown in Fig. 4. Here a large number of time transients were recorded for fixed values of ζ and the variable W. On the average approximately three runs were made at each setting. From each of these traces the time necessary for the system to reach the first point of zero error was recorded and plotted. The results are definite and indicate without question the increase in response of a system as W is decreased from linear operation toward a linear width of zero. The saturation effect of the curves is the result of W reaching the magnitude of the step input, and no further changes result, as would be expected. The effect of ζ is as would be anticipated, having little or no effect as $W \to 0$, but slowing the system down as ζ is increased for larger values of W.

Conclusions

It may be concluded from the experimental work that a definite improvement in response time may be obtained as the half linear width W is decreased toward true optimum relay operation. If perfect switching and complete compensation of the system could be attained, no overshoot would occur for the relay system. However, since neither of these factors were attained in the test system or would be realized in most applications, the increase in response time is attained at some sacrifice in overshoot performance. It must be kept in mind that the actual system is somewhat more damped than a true type 1 servo with equal values of ζ due to the uncompensated motor damping and the squared-error rate term not considered in the evaluation of ζ.

Further tests are underway to provide direct correlation between the calculated response times and the experimental performance. Work is also underway to adapt the system to direct analog computer operation such as the Reeves Electronic Analog Computer. It is felt that this may eliminate certain difficulties of stability and repetition accuracy found in the first computer set up.

Appendix

Integration of Equation 18

Equation 18 is given as

$$-J\ddot{E} = \frac{T_m}{W} \left[E + \frac{J}{2T_m} \dot{E}|\dot{E}| - W \right] \quad (18)$$

which becomes

$$-j\dot{E}\frac{d\dot{E}}{dE} = \frac{T_m}{W}E + \frac{J}{2W}\dot{E}|\dot{E}| - T_m \quad (21)$$

Consider the case where $\dot{E}$ is negative so that

$$\frac{J}{2W}\dot{E}^2 - J\dot{E}\frac{d\dot{E}}{dE} = \frac{T_m}{W}E - T_m \quad (42)$$

$$\text{Let } \dot{E}^2 = v \quad (43)$$

$$\frac{J}{2W}v - \frac{J}{2}\frac{dv}{dE} = \frac{T_m}{W}E - T_m \quad (44)$$

Rearranging

$$\frac{dv}{dE}-\frac{v}{W}=\frac{2T_m}{J}\left(1-\frac{E}{W}\right) \tag{45}$$

This is similar to Bernoulli's equation and has an integrating factor $e^{-E/W}$ so that

$$ve^{-E/W}=\frac{2T_m}{J}\int\left(1-\frac{E}{W}\right)e^{-E/W}dE \tag{46}$$

Performing the integration of equation 46,

$$ve^{E/W}=\frac{2T_m}{J}Ee^{-E/W}+K_1 \tag{47}$$

K_1 is a constant of integration. Hence

$$v=\frac{T_m}{JW}E+K_1e^{E/W} \tag{48}$$

or $$\dot{E}^2=\frac{T_m}{JW}E+K_1e^{E/W} \tag{19}$$

If $\dot{E}$ has been assumed positive, the result would have been

$$\dot{E}^2=-\frac{T_m}{JW}E+K_1'e^{-E/W} \tag{49}$$

References

1. The Application of Non-Linear Techniques to Servomechanisms, K. C. Mathews, R. C. Boe. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 8, 1952.

2. A Quasi-Linear Dual-Mode Servomechanism, P. P. Halverson. *M.S. Thesis*, Ohio State University, Columbus, Ohio, 1954.

3. A Nonlinear Servomechanism with Nonlinearities Resulting from Squaring and Saturation Functions, F. G. Kilmer. *Ibid.*, 1955.

4. Performance Characteristics of a Dual-Mode Servomechanism with Nonlinear Parameters, G. C. Milligan. *Ibid.*, 1956.

Discussion

Michael Athanassiades (University of California, Berkeley, Calif.): This paper represents a definite contribution towards the design of nonlinear control systems. The experimental results presented by the authors may be used directly for the design of a nonlinear system with the same equipment requirements.

The authors present in Fig. 4 the time required for the system in order to reach zero error versus the width of the linear zone W. In addition they point out that for small values of error rate the system behaves as a linear one, and natural frequency and damping ratio are given in equations 30 and 31. Since for small signals the system behaves as a linear one, then, theoretically, the time to reach zero error is infinite. The time indicated in Fig. 4 must be changed to indicate a settling time rather than time to reach zero error.

In the design of the nonlinear computer required, the authors have assumed an expression for the motor torque T given by equation 10. I should like to point out an alternative method of design yielding the same results, and using the phase plane exclusively. This method of design is based upon the Zero-Force-Curve (ZFC) concept. ZFC is the locus of all points in the phase plane for which the input to the saturating amplifier, and hence the motor torque, is zero. If the equation of ZFC is assumed in the phase plane, the switching boundaries and the system equations are easily determined.

Consider the normalized system of Fig. 5. ZFC is obtained by equating $\theta_e=0$. The $+1$ switching boundary is defined by $\theta_e=+1$ and the -1 switching boundary by equating $\theta_e=-1$. For large values of error let ZFC have the equation

$$\dot{E}^2=-\frac{\dot{E}}{|\dot{E}|}2E-2(1+A) \tag{50}$$

where A may be positive or negative. Then the $+1$ switching boundary has the equation for the lower part of the phase plane

$$\dot{E}^2=-\frac{\dot{E}}{|\dot{E}|}2E-2(1+A)-2 \tag{51}$$

and the -1 switching boundary the equation for the lower part of phase plane

$$\dot{E}^2=-\frac{\dot{E}}{|\dot{E}|}2E-2(1+A)+2 \tag{52}$$

Examining equation 50 we observe that for $\dot{E}=0$, $E=\pm(1+A)$ therefore the system will have an undesirable steady-state error of magnitude $1+A$. To avoid the steady-state error ZFC, for small error and error rate, must be gradually shifted so that it will pass through the origin.

The method of design is as follows:

1. Choose a particular value of $\dot{E}=\dot{E}_m$. For this value of error rate ZFC defines a point E_{om} on the error axis

$$E_{om}=-\frac{\dot{E}}{|\dot{E}|}\left[1+A+\frac{\dot{E}_m^2}{2}\right] \tag{53}$$

2. Change the equation of ZFC in the region $\dot{E}_m\leq E\leq -\dot{E}_m$ or region 3 of Figs. 6, 7, and 8, so that it will pass through the origin of the phase plane and the points

$$\left\{-(1+A)-\frac{\dot{E}_m^2}{2},\ \dot{E}_m\right\} \text{ and } \left\{(1+A)+\frac{\dot{E}_m^2}{2},\ -\dot{E}_m\right\}$$

Fig. 6 shows the phase-plane boundary structure with ZFC being a straight line in region 3. This system will be very similar to the one described by the authors for the particular value of $A=0$.

Fig. 7 represents a system with a region 3 ZFC given by a parabola with vertex at the origin and passing through E_{om}, $-\dot{E}_m$ and the points $-E_{om}$, $\dot{E}_m$. This system will be faster than that of Fig. 6, but will have more overshoot, because the slope of ZFC at the origin is infinite.

Fig. 8 represents a system with a region 3 ZFC given by a parabola with vertex at the point $\pm E_{om}$ $\mp\dot{E}_m$ and crossing the origin of the phase plane with a finite slope. This system will be slower than those previously described, but will have negligible overshoot.

Of course there are an infinite number of region 3 ZFC. They may be of any shape, and the only requirement is that they pass through the origin of the phase plane and the points $(\pm E_{om}, \mp\dot{E}_m)$. The ones previously described give system equations of moderate complexity, and the nonlinear computer may be built to reproduce the nonlinearities required without nonlinear networks.

For a given set of specifications and a set of experimentally obtained curves indicating rise time, peak overshoot ratio, and settling time, as functions of the magnitude of the step input, and the variation of parameters A and $\dot{E}_m$, an optimum design may be obtained. The method of obtaining the optimum design is based upon the concept of error measure.

For example an error-measure function may have the form

$$\text{Error Measure}=R(\text{Rise time})^p+P(\text{Peak overshoot ratio})^p+S(\text{Settling time})^p$$

where the coefficients and exponents depend upon the relative importance of the transient quantities.

If the frequency of occurrence of the step inputs is known as a function of their magnitudes, a probability density curve may be obtained for the input. The optimum design is the one that minimizes the error measure, and may be obtained using graphical methods.

If the set of specifications include the requirement of finite settling time to zero error and zero overshoot, the design of Fig. 9 may be used. The unique characteristic of this design is that ZFC and the switching boundary intersect at the origin of the phase plane. The design is symmetrical about the origin.

The equations of ZFC and the switching boundaries are

Positive Error Region

-1 Switching Boundary.... $\dot{E}^2=2E$

ZFC..... $\dot{E}^2=\dfrac{2E^2}{1+E}$

$+1$ Switching Boundary.... $\dot{E}^2=\dfrac{(E-1)2E}{1+E}$

Negative Error Region

-1 Switching Boundary.... $\dot{E}^2=\dfrac{(E+1)2E}{1-E}$

ZFC..... $\dot{E}^2=\dfrac{2E^2}{1-E}$

$+1$ Switching Boundary.... $\dot{E}^2=-2E$

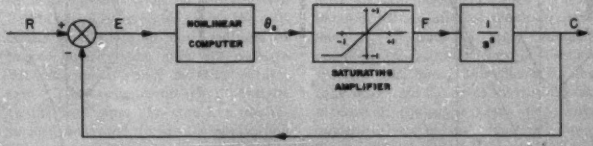


Fig. 5. Normalized system

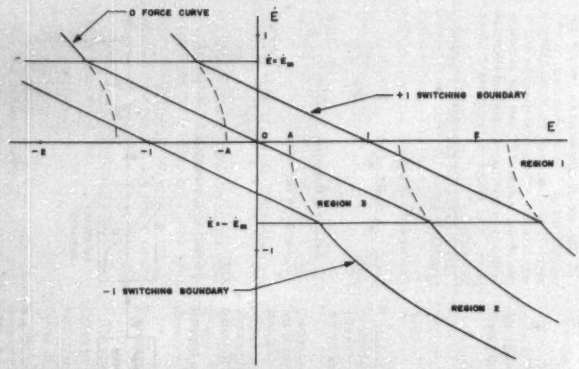


Fig. 6. Phase-plane boundary structure with ZFC as a straight line in region 3

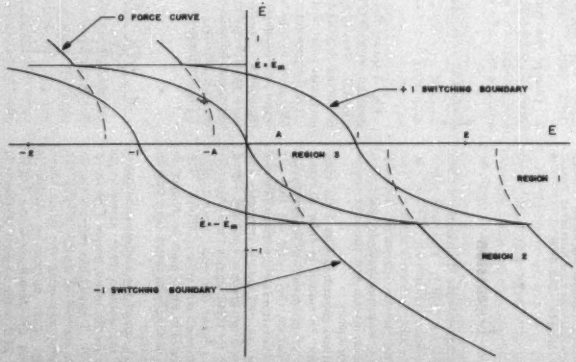


Fig. 7. System with region 3 ZFC given by a parabola with vertex at the origin and passing through E_{m_0}, $-\dot{E}_m$ and the point $-E_{m_0}$, $\dot{E}_m$

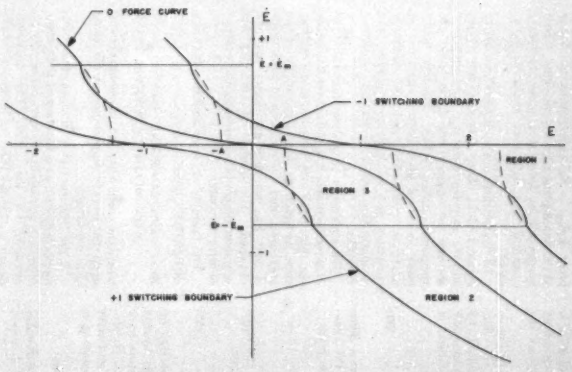


Fig. 8. System with region 3 ZFC given by a parabola with vertex at the points $\pm E_{m_0}$, $\mp \dot{E}_m$ and crossing origin of phase plane with finite slope

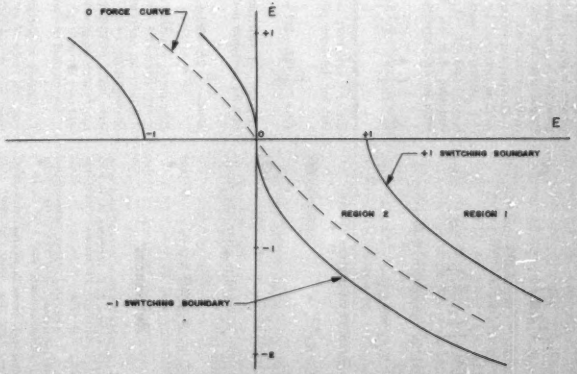


Fig. 9. System with finite settling time to zero error and zero overshoot. ZFC and switching boundary intersect at the origin of the phase plane

The input to the saturating amplifier is given by

$$\theta_c = E - \left[\frac{1}{2E} + \frac{E}{|E|}\frac{1}{2}\right]\dot{E}^2$$

In conclusion the ZFC method of design on the phase plane offers the advantage of visualizing the small signal behavior of the system better than the mathematical formulation. The large signal behavior of the system remains virtually unchanged for the various small signal design methods.

H. R. Weed and F. C. Weimer: This discussion by Mr. Athanassiades is appreciated and of interest in the field. With respect to the specific comment concerning the co-ordinates of Fig. 4, the time to reach zero error was actually the time to the first zero crossing of the transient response of the oscillatory system. This time was used because it was easily read from recorder data. However, as Mr. Athanassiades suggests, a settling time would be better, although somewhat more difficult to obtain. Complete experimental data will be discussed in a future paper. The zero-force-curve concept introduced by Mr. Athanassiades is a different method of attack, which should be useful in design.

Static Control in Automatic Warehousing

L. L. BOSCH
FELLOW AIEE

A. J. FANTHORP
NONMEMBER AIEE

J. W. STUART
MEMBER AIEE

THE virtual elimination of relay contacts from the highly repetitive functions required for automatic warehousing was made possible by the selection of static control elements, Westinghouse Cypak,[1,2,3] in place of conventional relays for an extensive mechanized carton-handling installation in the Chelsea Milling Company premix food plant at Chelsea, Mich. While the basic concept of the control philosophy was determined to be the same for either relays or static control elements, the high reliability and negligible maintenance inherent in static control elements, which have no moving parts subject to wear or corrosion, dictated the choice. Interestingly, the first cost was found essentially comparable.

The logic function approach is used in Cypak circuit design in which, as work in the communication field has shown, any set of conditions or arrangements, no matter how complex, can be described in terms of four functions, the AND, OR, NOT, and MEMORY. Application of these fundamentals is employed in the makeup of the Cypak AND (substitute for relay contacts in series), the OR (substitute for relay contacts in parallel), the NOT (substitute for normally closed relay contact), and the MEMORY (substitute for the "latched-in" relay) functions. These four basic elements are the building blocks for Cypak control as will be noted by reference to Table I, which shows the relay equivalents and operation for a representative number of Cypak elements.

The operation of all Cypak elements, except the OR, is based on a magnetic amplifier principle coupled with a Ramey amplifier circuit. Special Hipernik-type steel is used to form the amplifier core. This metal is an alloy (approximately 50% nickel, 50% iron) providing an approximately square hysteresis loop which is ideal for self-saturating operation. Using this magnetic material as a core and applying an alternating voltage to a winding (gating) causes an exciting current to flow through the winding. With proper values of voltage and the proper number of turns in the winding, the resulting current will drive the core to saturation on the positive half-cycle voltage. A rectifier, placed in series with the gating winding, prevents flow of current on the negative half-cycle of voltage; and the core remains saturated. On the next positive half-cycle, since the core cannot absorb more flux, voltage will appear across any load connected to the amplifier.

A second winding (reset) is added to the magnetic core so that when a separate voltage is applied, it will drive the core to negative saturation on the negative half of the voltage cycle. With the continued application of both voltages of proper polarity, the core will oscillate between positive and negative saturation. There will be no output voltage to the load since the core has not remained saturated during the application of a second one-half cycle of gating voltage. If the reset voltage is absent for one reset half-cycle, an output will appear across the load resistor on the next gating half-cycle. The input or controlling voltage to the Cypak is applied to the reset winding and associated bias resistors and rectifiers in a manner which operates to permit or not permit the reset voltage to cause negative saturation of the magnetic core.

An understanding of the control circuits, which accomplish the automatic warehousing, will be facilitated by first referring to Fig. 1, a pictorial representation of the mechanical handling system. It will be noted that the conveying system starts at the respective carton sealer discharges in the cake, biscuit, and pie crust Jiffy product mixing departments. Memory circuits between these departments and the finished goods warehouse cause the storing of information giving the sequence in which the respective product cartons enter the conveying system. From the three production areas, the cartons are conveyed to merging points and then single file, by means of powered and gravity conveyors, to a jib crane located in the finished goods warehouse some several hundred feet distant. Here the cartons are sorted and dispatched automatically to preselected magazine or on-storage conveyor lines; see Fig. 2. This paper is concerned particularly with a description of the control methods used in accomplishing the automation.

Control Block Diagram

A review of the over-all block diagram, Fig. 3, will show that two sets of paper-tape punch and reader equipment are used for the memory function. Initial impulses for operation of the first punch, line A, originate in relay contacts operated by gate switches in the respective cake lines. Ingredients for the cake mixes are fed from overhead food-processing equipment into five fully automatic packaging machine lines. These fill and seal individual retail-size packages with the various food mixes,

Paper 58-1285, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted October 25, 1956; made available for printing September 10, 1958.

L. L. BOSCH and A. J. FANTHORP are with Bosch & LaTour, Cincinnati, Ohio, and J. W. STUART, formerly with Westinghouse Electric Corporation, Pittsburgh, Pa., is with the Industrial Nucleonics Corporation, Columbus, Ohio.

Table I. Cypak Appendix

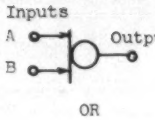
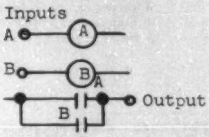
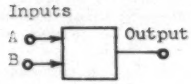
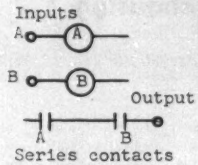
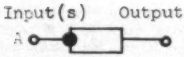
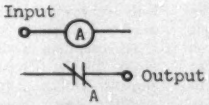
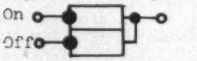
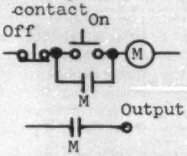
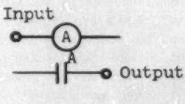
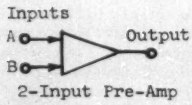
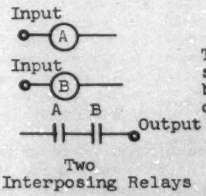
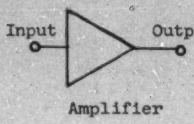
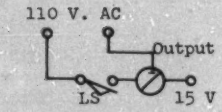
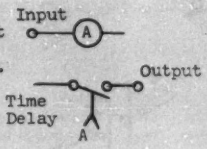
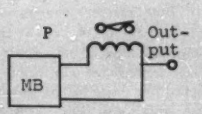
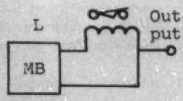
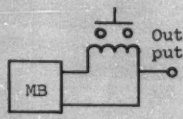
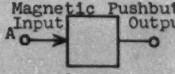
SYMBOL	OPERATION	RELAY EQUIVALENT	REMARKS
 OR	A input OR B input provides an output.	 Parallel contacts	OR prevents "back-feed" into either input--contacts do not.
 2 Input And	A input AND B input must be present to obtain an output.	 Series contacts	All CYPAK units provide several outputs as required.
 Not	With input A present an output can not be obtained. With an input not present an output results.	 Normally closed contact	
 Memory (2 Nots)	Memory remembers the most recent input. It provides an On output voltage and a distinct Off output voltage, but with the On and Off inputs applied simultaneously it has no output voltage.	 Sealed circuit	
1-Input Pre-Amp	With an input present an output is obtained.	 Interposing Relay	This CYPAK unit is same as 1 Input And but with higher wattage output capacity.
 2-Input Pre-Amp	Both inputs (A and B) must be present to obtain an output.	 Two Interposing Relays	This CYPAK unit is same as 2 Input And but with higher wattage output capacity.

Table I. (cont.)

 Amplifier	With an input present the amplifier provides an input to contactor coils or valve solenoids.	No equivalent.	
 Input Transformer		No equivalent.	Matches high voltage contact requirements with low voltage CYPAK.
Delay Element	With an input present full output will be developed after a time delay.		Timing element is used with Memory for snap action after timing period elapses.
 Magnetic Limit Switch (P) Proximity Type	A normally closed LS which will open when a soft iron bar comes in proximity with the switch.		Contactless Limit Switch.
 Magnetic Limit Switch (L) Lever Type	A normally closed LS which will open when the lever is operated.		Contactless Limit Switch.
 Magnetic Pushbutton			Contactless Pushbutton.
 1 Input And	With A input present an output is obtained.	No equivalent.	Used as buffer amplifier and phase shifter.

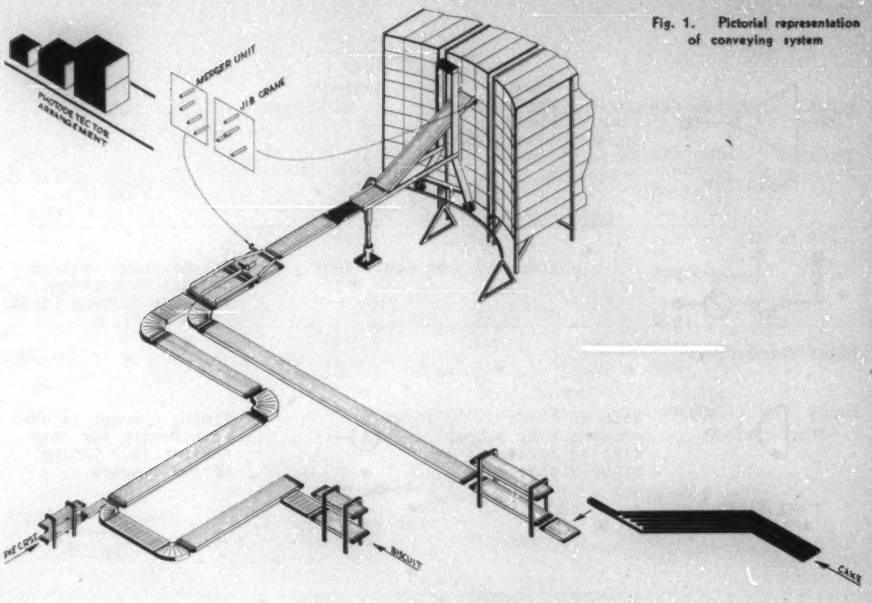


Fig. 1. Pictorial representation of conveying system

such as white cake, yellow cake, and chocolate fudge cake.

Lines of sealed cake packages cumulate on discharge aprons at each machine where they are held momentarily by a remotely controlled gate, adjustable for release of the exact multiple of packages needed to fill one shipping carton—in this instance, six. The packages are stuffed into the cartons by essentially automatic equipment and after passing through carton sealer belts, enter line *A* conveyor.

As shown in the block diagram, circuits from the respective lines are connected to plug cords marked *L-1*, *2*, *3*, *4*, and *5*. The plugs in turn are inserted manually into jacks corresponding to the products being made on the five lines. Twenty-three jacks are provided for cake mix products. When a train of packages is released, one of the line limit switches, such as *LS-1*, is closed. This action, as will be seen later, also activates an impulse generator. This creates one pulse for each carton that is to be filled.

The impulse generator, Fig. 4, consists of an aluminum disk which is driven continuously by a 110-volt a-c motor. The disk contains one cast-iron slug, which operates a start–stop magnetic proximity switch, and positions for six similar slugs about its periphery which act to generate pulses as the slugs pass a second proximity switch. The number of slugs to be mounted corresponds to the number of carton units scheduled for release from the stuffer.

Closing a package line gate switch allows the start–stop proximity switch to turn on a memory. As is to be noted in Fig. 4, MEMORY-*2* and the output of the impulse proximity switch through MEMORY-*1* provide the inputs to AND-*2*, which in turn provides the output to the coding ANDs controlling punch no. *1*. This is applied to a diode matrix through the Cypak plugboard, Fig. 5, which translates the 23 decimal product numbers into 5-digit binary numbers for energizing the required combination of magnetic amplifiers to provide intelligence to the punch magnets and index clutch coil of tape no. *1*. This provides storage information (MEMORY) for subsequent use.

All punch and reader equipment is

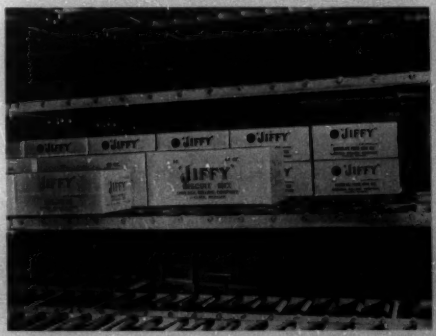


Fig. 2. Carton sizes and arrangement handled on system

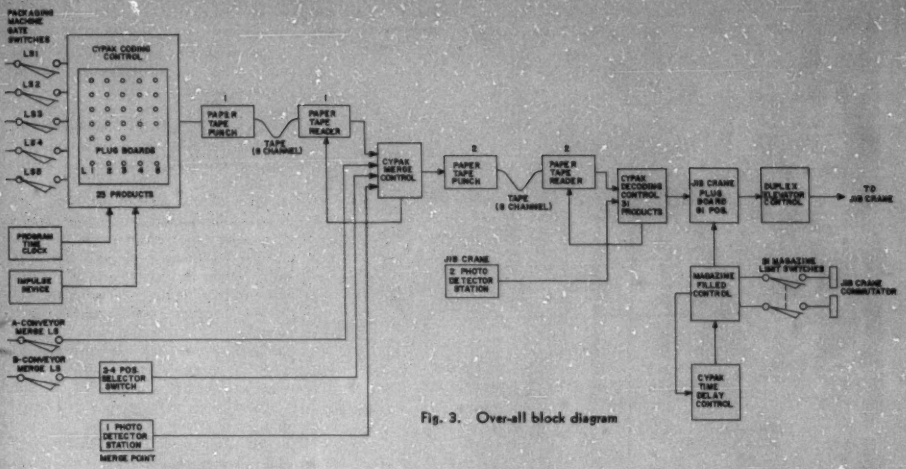


Fig. 3. Over-all block diagram

located at the jib crane and the tape has eight channels: five for product identification and three for time signal from program clock and for error coding. The punch and reader assemblies are modifications of standard motorized units of the Commercial Controls Corporation. Paper tape from each punch is fed directly into the adjacent paper-tape reader equipment and each pair of units is mounted to provide a minimum span of tape.

The reader of punch no. *1* conveyor line *A* is actuated by phototransistor detection equipment and switches located at the merge point of conveyors *A* and *B*.

The conveyor equipment for merging is known as a Merg-A-Flow unit, supplied by Alvey-Ferguson Company, manufacturers of the other conveying equipment and the jib crane. It consists of a double-width belt for cumulating cartons from lines *A* and *B*, an automatically controlled stop arm in each line, and a discharge belt operating at higher speed to get separation for photodetection. Photodetection (Fig. 6) accomplishes two results:

1. It sets up a circuit in the Cypak merge control. This reads the characters on paper tape no. *1* each time a carton from line *A* passes the no. *1* station. It also energizes the reader clutch which indexes the tape for the next carton. The coded information, read from paper tape no. *1*, is immediately retransmitted to paper tape punch no. *2* on which the identical respective characters of tape no. *1* are punched. This operation working through a 3-input AND circuit (line switch,

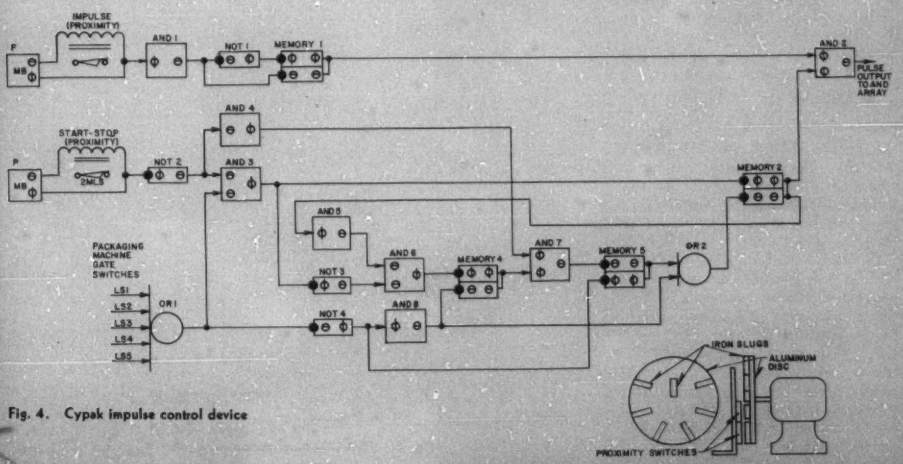


Fig. 4. Cypak impulse control device

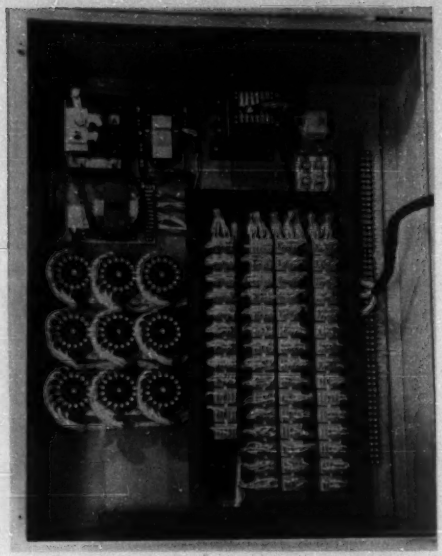


Fig. 5. Cake mix department Cypak panel

reader no. *1*, and phototransistor) continues as long as conveyor line *A* is released. 2. It differentiates between the two carton sizes originating on line *B* (large in biscuit department, cake size in pie crust department). When line *B* is released it causes preselected characters to be punched on paper tape no. *2*. These characters are determined by the settings of two 4-position selector switches, one in the biscuit and the other in the pie crust department. Each of the 4-position switches corresponds electrically to the plugboard in the cake mix department.

In operation, line *A* is the normally released line. Line *B* is released for an adjustable time interval when a predetermined number of cartons accumulates on line *B* or when a tight-tape signal is received from tape reader no. *1*. This latter signal indicates that, at the moment, the in-transit production from line *A* is insufficient to create the required length of slack tape between the no. *1* punch and reader.

During the period when line *B* is released, the punching of tape no. *2* is determined from the output of either of two 2-input ANDs. One corresponds to the pie crust line and one to the biscuit line. A common input from the *B* line release switch enters each of these selected ANDs, while a second input enters only the AND unit of the packaging line, producing the box size detected by the phototransistor unit.

It will be noted that the tape at punch no. *2* records the sequence of all cartons passing the *A–B* merging station regardless of where the cartons originate. Carton flow is single file from the merging station to the jib crane. Reader no. *2* is read each time a carton or carton unit is detected at the tip of the jib crane conveyor.

The respective signal voltages from tape reader no. *2* are decoded in the Cypak decoding control by an AND matrix. This action translates the 5-digit binary product numbers into decimal number identification and, in turn, operates the proper product identification relay; see Fig. 7. The outputs of these 31 distinct product relays are then fed into a corresponding number of plug cords on the jib crane dispatch board. Each plug is physically arranged on reel equipment, so it can be inserted into any one of the 9-wide 9-high conveyor magazine jack locations. Each jack location forms two definite crossbar (bus) connections not common to any other magazine jack. Consequently, when a product plug is inserted into one of these jacks, each product has a definite crossbar connection not common to any other product. Ordinarily, only plugs corresponding to products in actual or pending production would be inserted into jacks.

The two voltages provided by the crossbar circuit connections result in one to nine floor calls to each section of duplex elevator controls (manufactured by Shepard-Warner Company): one, the vertical, corresponding to the proper vertical conveyor line, and the other, horizontal, corresponding to the proper conveyor row.

The jib crane, as will be noted in Fig. 1, has a conveyor belt mounted

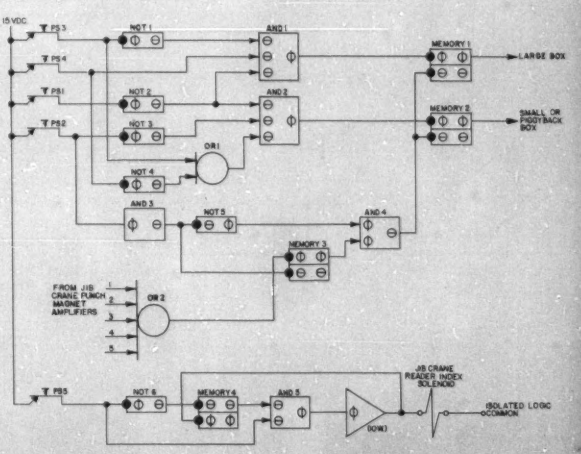


Fig. 6. Photodetection station schematic

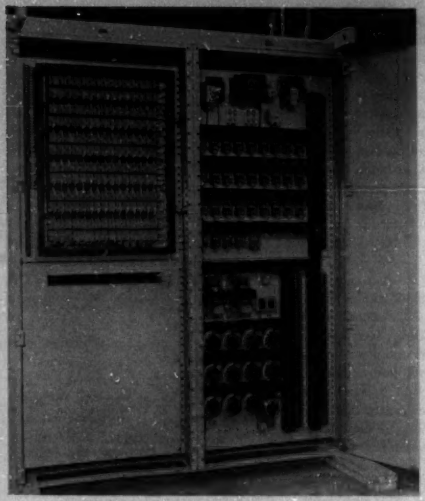


Fig. 7. Jib crane Cypak panel

atop the boom. This belt is divided into two sections, a short rotatable horizontal piece which is immediately above the crane's pivot point and a long hinged length which permits the vertical motion. An alligator connection is made between the incoming fixed belt and the rotatable section. A single motor drives both the horizontal and the hinged section but at differential speeds to accomplish carton spacing. Its control is interlocked so as to stop a brake-type supply-line belt whenever the belt of the crane is stopped which occurs whenever the crane repositions itself. A live roller section of conveyor is located immediately upstream of the brake belt. This is for accumulating cartons when the crane is in motion. Two completely independent, 2-speed reversing-type drives are used for controlling the position of the jib crane; one is actuated by the vertical automatic elevator controller, and the other, by the horizontal.

The vertical movement is by a winch cable drive, while the horizontal is by a geared truck mounted on a monorail track.

While, as indicated in Fig. 1, a number of phototransistors are mounted near the end of the boom, only one is actually used in the performance of the dispatching operation. The others are for self-determining error checks.

When a carton interrupts the light to the controlling phototransistor, it sets in operation a circuit which triple indexes the tape to the next set of characters. These are then read by the Cypak circuit for product identification as previously described and the result referred to the dispatch board for signaling simultaneously to the two sets of automatic elevator control for jib crane positioning. This entire operation and computation requires slightly less than one-half second. If, when the signals are received by the elevator controllers, it is determined that the crane is already in its proper horizontal and vertical position for discharging the carton, the motor driving the boom belt is permitted to continue to run; if not, it is stopped while the crane repositions itself automatically. As soon as this is accomplished, the belt is started again for dispatch of the next carton to its proper location.

As a carton enters a given magazine conveyor, a limit switch, located about 8 feet downstream, is momentarily tripped. This initiates a time-delay unit in the Cypak control panel which, if timed out, causes operation of an all-filled conveyor relay in the jib crane plugboard. The relay is of the latched-in type which, when operated, acts to transfer all subsequent elevator control calls to the next choice location as predetermined by jack-to-jack jumper cords inserted into the dispatch board. When the magazines have been unloaded, the original sequence is restored by push-button release of the latched-in holding coils. The employment of a single Cypak time-delay operation for all positions is made possible through the use of elevator-type commutators mechanically tied to the jib crane movements. Because of the short time interval for timing circuit reset between cartons, it was necessary to use a flip-flop type of Cypak timing circuit.

Each time a biscuit-size carton is detected at the tip of the jib crane, an error circuit is energized which checks to determine that the signal obtained from the memory circuit is for a biscuit-size carton. If not, the belt is stopped automatically and a horn signal given to attract a qualified man to check on the error. The same operation occurs on the arrival of a carton from the cake department which, for signaling or other reason, had an error signal imposed on its product code. This signal is impressed by the manual operation of a control switch on the cake department Cypak panel.

While, as will be noted in Fig. 1, the conveyor storage racks are nine high and nine wide, only 72 of the theoretical 81 positions are used. Each rack is a roller-type gravity conveyor up to 200 feet in length on which the cartons are temporarily stored. For accessibility, the 72 on-conveyor storage lines are contained in three magazine sections. These are supported by a Powerstrut mechanical support system in a 3-wide, 7-, 8-, or 9-high honeycomb arrangement.

In order to increase the potential of the storage system, facilities have been provided for "piggy-backing" all of the cake-size cartons. Since the ratio of the number of these sizes to the biscuit size is large, the storage-capacity increase is substantial. The mechanisms which accomplish the piggy-back on the respective lines were especially built. These are located immediately downstream of the carton sealers. The electric control for piggy-back is the same as for single carton units except that only three, rather than six, slugs are used in the impulse generator. This creates one half the number of codes. The aggregate of roller conveyor storage totals about 11,300 lineal feet which, with piggy-back, is nearly equal to the total of the second- and third-shift production.

A number of magazine conveyor switch-out sections are available to permit shunting to on-floor stacked storage of production that is not scheduled for immediate shipment. The cartons are held in the magazine by brakes located at the end of each conveyor. These are manually operated by small

crank arms which can be adjusted from the full stop to any throttled discharge position. Discharge of the on-storage conveyor lines occurs during the first and part of the second shift. Within the limits of practicability, this operation is scheduled so the conveyor system is empty of both stored and first-shift production by the beginning of the second shift.

A simple weighted chute at the end of each on-storage conveyor permits directing the released cartons onto a 3-wide, inclined, powered conveyor. Each of the three inclined conveyors is connected through power or gravity conveyor lines to a 4-high swing belt. Three such swing belts are grouped into one assembly. These belts, in turn, are connected to other conveyors which lead to three rail docks and one combination truck-rail dock. This arrangement permits delivery of cartons from any of the 69 conveyor lines to any loading dock without manual handling. In actual operation, loading may vary from a solid railroad car of one product to a truck having a wide variety of products arranged for store-to-store route deliveries.

Other plans and arrangements were studied before the afore-described system was selected, but each of the alternatives lacked the flexibility and economy required to accomplish the desired result. One of the by-products and significant long-term benefits anticipated in the selected system is the saving in carton breakage. Throughout the planning, emphasis has been laid on reducing material handling costs which are generally estimated as 25–30% of all direct manufacturing costs.[4]

Conclusions

This food product installation indicates that, with the dependable control methods now available, automation will undoubtedly be extended into more process plans than were heretofore considered practical. It shows that by combining commercially developed control devices, a very special tailor-made result can be obtained. A paper punch tape has been used as a memory circuit; automatic elevator control for jib crane operation; and Cypak static control units for coding, decoding, and other operations. The use of Cypak has simplified the design and resulted in a system which should require minimum maintenance over the years. The end result of this installation is a conveyor system which sorts out 31 products and dispatches them to 72 on-storage conveyor lines having an aggregate of 11,300 feet of lineal carton storage space. The system makes possible shipment of three shifts of daily production in an essential single shift without manual handling from production to the shipping facility. Further, it reduces to a minimum the manual warehouse handling of those cartons which are not scheduled for immediate shipment. The tape record is available for production analysis.

References

1. On the Control of Magnetic Amplifiers, R. A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2124–28.

2. CYPAK Systems, Editorial Staff. *Westinghouse Engineer*, East Pittsburgh, Pa., July 1955, p. 4.

3. CYPAK Systems, J. P. Baker. *Ibid.*, July 1956, p. 2.

4. Plant Layout (book), R. W. Mallick, A. T. Gaudreau. John Wiley & Sons, Inc., New York, N. Y., 1951.

General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems (An Optimal Sampling System)

R. E. KALMAN
NONMEMBER AIEE

J. E. BERTRAM
ASSOCIATE MEMBER AIEE

Synopsis: This paper is concerned with the problem of designing optimal systems for the control of a plant governed by a linear differential equation with constant coefficients. Control is exerted by means of piecewise constant signals which can change only at the "sampling instants." Optimality means here, as in some nonlinear problems, that the system achieves equilibrium with zero steady-state error from any initial state as quickly as possible.

The sequence of signals required for optimal control in this sense is ideally a linear combination of the state variables of the plant being controlled. In the simplest regulator problem, when all state variables are accessible, the optimal system is realized by a very simple multifeedback arrangement in which the only unconventional component is the sample-and-hold element. When some states are not accessible, when the plant includes time delays, and in some types of follow-up systems, the physical realization of optimal control requires real-time analog or digital computation. The requirements of machine computation for control purposes are derived here in a general way; the theory presented includes as special cases many commonly used methods for control system synthesis.

IT HAS BEEN pointed out in several recent papers[1–4] that the use of sampled-data techniques results in control systems superior in many respects to unsampled systems. Yet until now few practical applications of sampled-data control systems exist, except in those cases where the input or error signal of the system is inherently sampled. Probably the major reasons why sampling techniques have not been more widely used is the complexity and expense of the sampling controller compared to the simple elements available to compensate unsampled systems. Further, in those cases where sampling controllers have been tried, it was found that adjustment of the controller parameters for optimal operation would be very difficult unless the plant transfer function could be measured with high accuracy. An additional difficulty is that a sampling system de-

Paper 58-1161, recommended by the AIER Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted March 13, 1958; made available for printing July 25, 1958.

R. E. Kalman is with the Research Institute for Advanced Studies (RIAS), Baltimore, Md., and J. E. Bertram is with International Business Machines Corporation, Ossining, N. Y.

This research was supported in part by the U.S. Air Force under Contract No. *AF 18(600)-477* at Columbia University, New York, N. Y., and was monitored by the Office of Scientific Research, Air Research and Development Command.

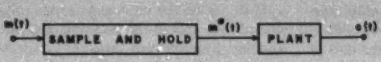


Fig. 1 (left). The system to be controlled

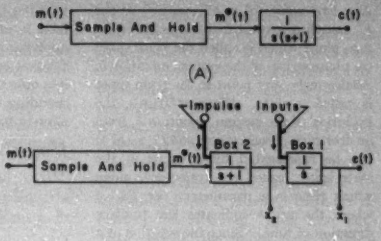


Fig. 2 (right). (A) An example. (B) Selection of state variables

signed by conventional methods to perform in an optimum fashion at sampling instants does not always have satisfactory performance at other times.[5]

In this paper a design technique is developed which results in optimal performance at all times, not just at the sampling instants. The sampling controller for optimal performance can be easily constructed. The experimental adjustment procedure to achieve optimal performance is found to be quite simple.

The significance of the principles introduced here is quite broad; numerous additional matters such as compensation for time delay and saturation, proper use of real-time machine computation, etc., are discussed briefly. This paper is directly mainly to questions of engineering interest; a rigorous and detailed discussion of the mathematical questions will be found in an (as yet) unpublished paper by R. E. Kalman, entitled, "On the General Theory of Control Systems."

Statement of Problem

Consider the system shown in Fig. 1 which consists of the following.

1. The plant, which is governed by an ordinary, linear, time-invariant differential equation. As is customary in the control systems literature, the plant is described by a rational transfer function.

2. The sample-and-hold element, which replaces the continuous control signal $m(t)$ by a piecewise constant or sampled signal $m^*(t)$, described by

$$m^*(t)=m(kT),\ kT\leq t<(k+1)T,\quad k=0,\ 1,\ 2,\ldots \qquad (1)$$

where the sampling period T is a positive constant, the point $t=kT$ in time is called the k^{th} sampling instant. Frequently, it will be convenient to write $f(k)$ instead of $f(kT)$, where $f(t)$ is some function of time.

The design of control system can be formulated in general as follows: Given an arbitrary input signal $r(t)$, select the control signal $m(t)$ in such a way that the error $e(t)$

$$e(t)=r(t)-c(t) \qquad (2)$$

is as small as possible in some precisely defined sense.

This paper is concerned primarily with the following problem:

Find the function of time $m(t)$, that is, the sequence of numbers $m(0)$, $m(1)$, ... which will bring the plant from any condition existing at $t=0$ to equilibrium with zero error in a minimum number of sampling periods.

In the following sections a complete solution of this problem will be given in the special case when $r(t)=$constant. Such a control system is called a regulator; the problem stated is then that of designing a regulator which eliminates the effect of any suddenly applied disturbance to the plant in the minimum length of time consistent with a fixed choice of the sampling period T.

If $r(t)$ is an arbitrary function of time, the problem cannot be solved in general. However, when $r(t)$ is of the form

$$r(t)=\sum_{i=1}^{m} b_i t^{i-1},\ t\geq 0 \qquad (3)$$

$$=0,\ t<0$$

(where the b_i are arbitrary real numbers) the solution can be readily obtained by modifying the solution of the regulator problem.

Conventional methods[2] for designing sampled-data systems are frequently derived in such a way as to make $e(kT)=0$ at every sampling instant after, say N sampling periods following the application of a prototype input (such as a step or a ramp) to a system at rest. The disadvantage of such methods is that the plant may not have reached equilibrium in N steps, so that even though the error will be zero at sampling instants it will not be zero between the sampling instants at any time $t\geq NT$. This phenomenon is well known in the theory of sampled-data systems and is referred to as intersample "ripple." In the present formulation of the problem, the plant is required to reach equilibrium when the error becomes zero. As a result, the ripple problem is completely eliminated.

It is interesting to note, as a matter of historical interest, that the present formulation of the problem, based on the ripple-free requirement, was announced by one of the present authors in 1954, together with the solution in a simple special case.[6] At that time, the basic idea was derived from certain well-known methods in the theory of nonlinear (saturating) control systems; here the earlier idea is generalized using techniques for the theory of nonlinear dynamic systems.

System Description

At the sampling instants the behavior of the system shown in Fig. 1 may be represented by the equation

$$\mathbf{x}(k+1)=G(T)\mathbf{x}(k)+m(k)\mathbf{h}(T) \qquad (4)$$

where $\mathbf{x}(k)$ is a vector representing the state of the plant at the k^{th} sampling instant; $G(T)$ is an $n\times n$ matrix whose elements are functions of T, called the transition matrix[7] of the differential equation governing the behavior of the plant; $\mathbf{h}(T)$ is a vector which represents the effect on the state of the plant at the $(k+1)^{\text{th}}$ sampling instant of a unit-step control signal applied to the plant at the k^{th} sampling instant.

To illustrate how equation 4 is obtained explicitly, consider the example of Fig. 2(A). This example will be used throughout the paper. The first step is to redraw the block diagram of the plant in such a way as to make the state variables (i.e., the components of the state vector) accessible. This can be done by decomposing the plant into a larger number of block in such a way that each block contains a transfer function with one pole and no finite zero. There are many possible ways of doing this; however, the choice is considerably narrowed by requiring that as many of the state variables as possible must be physically measurable quantities. A possible choice is shown in Fig. 2(B).

To define and motivate the preceding terminology, consider now the differential equations corresponding to Fig. 2(B).

$$dx_1/dt=x_2$$

$$dx_2/dt=-x_2+m^*(t) \qquad (5)$$

Any two values of $x_1(kT)$, $x_2(kT)$ at the k^{th} sampling instant define a point in the x_1,x_2 plane which is called the state space (or phase space) of the system, equation 5. Analogously, any point in the state space is called a state. When $m^*(t)=0$, the evolution of the system, equation 5, from an arbitrary present state $(x_1(kT), x_2(kT))$ can be represented by a curve in the state space, called the trajectory, along which time is a parameter; see Fig. 3 where the arrow indicates the positive direction of time. Since the solution of a linear differential equation is uniquely determined for all finite values (positive or negative) of time by the initial state, there is one and only one trajectory passing through any given point in the x_1, x_2 plane. It is convenient to think of the behavior of the system, equations 5, in the absence of an input as the motion of a point representing the state along a trajectory. If the point does not move at all, i.e., if the trajectory through the point is the point itself, the point is called an equilibrium state. The equilibrium states of equation 5 are the states for which $dx_1/dt=dx_2/dt=0$.

The fundamental matrix expresses the dependence of any future (or past) state on the present state. To find the elements of this matrix, it is only necessary to express the state of the plant at any time explicitly as a function of the present state. Now, the present value $x_1(kT)$ can be brought about by applying an impulse at $t=kT$ of area $x_1(kT)$ to the input of box 1, if the system in Fig. 2(B) is initially in the zero state. A typical germ $g_{ij}(T)$ of $G(T)$ is therefore obtained by calculating the response of the plant observed at state variable x_i at time $t=(k+1)T$ when a unit impulse is applied at $t=kT$ to the input of the box containing state variable x_j. In case of Fig. 2(B),

$$G(T)=\begin{Vmatrix} g_{11}(T) & g_{12}(T) \\ g_{21}(T) & g_{22}(T) \end{Vmatrix}=\begin{Vmatrix} 1 & 1-e^{-T} \\ 0 & e^{-T} \end{Vmatrix} \quad (6)$$

To calculate the effect of the control signal $m^*(t)$, note first that it is constant between sampling instants by virtue of its definition. In other words, the control signal is always a step function. To obtain the response observed at state variable x_i at time $t=(k+1)T$ as a result of applying a step at $t=kT$ to the input of the box containing the state variable x_j, it is only necessary to integrate the corresponding impulse responses, which have already been determined. Thus:

$$\int_0^T G(\tau)d\tau=H(T)=\begin{Vmatrix} \int_0^T g_{11}(\tau)d\tau & \int_0^T g_{12}(\tau)d\tau \\ \int_0^T g_{21}(\tau)d\tau & \int_0^T g_{22}(\tau)d\tau \end{Vmatrix}=\begin{Vmatrix} T & T-1+e^{-T} \\ 0 & 1-e^{-T} \end{Vmatrix} \quad (7)$$

In the present example, however, the control signal is applied only to the box containing x_2. Therefore only the second column of $H(T)$ is of interest, so that

$$\mathbf{h}(T)=\begin{Vmatrix} T-1+e^{-T} \\ 1-e^{-T} \end{Vmatrix} \quad (8)$$

More generally, one can write

$$\mathbf{h}(T)=H(T)\mathbf{b} \quad (9)$$

where the vector $\mathbf{b}$ is the state resulting from the application of an impulse to the input of the plant which is initially in the state $\mathbf{x}=0$.

In explicit notation, the vector equation 4, for the example is:

$$x_1(k+1)=x_1(k)+(1-e^{-T})x_2(k)+(T-1+e^{-T})m(k)$$

$$x_2(k+1)=e^{-T}x_2(k)+(1-e^{-T})m(k) \quad (4A)$$

Two important properties of the transition matrix are needed in the sequel. First, if $m(k)=0$ for all k, then using equation 4 several times shows that

$$\mathbf{x}(k+l)=G(lT)\mathbf{x}(k) \quad (10)$$

for any integers (in fact, any real numbers) k, l. But, since by the afore-mentioned observation $\mathbf{x}(k)=G(kT)\mathbf{x}(0)$ and $\mathbf{x}(k+l)=G((k+l)T)\mathbf{x}(0)$, it follows that $\mathbf{x}(k+l)=G(lT)\mathbf{x}(k)=G(lT)G(kT)\mathbf{x}(0)=G((k+l)T)x(0)$ for any state $x(0)$. Therefore

$$G((k+l)T)=G(kT)G(lT) \text{ (any } k, l) \quad (11)$$

In particular, letting $l=-k$, it follows:

$$G^{-1}(kT)=G(-kT) \quad (12)$$

Equations 11 and 12 make it very easy to calculate powers and inverse powers of the transition matrix.

The Canonic Co-ordinate System

Consider now the solution of the basic problem, with $r(t)=0$. In other words, it is desired to find a sequence $m(0)$, $m(1)$, ... which will take any initial state $\mathbf{x}(0)$ to equilibrium with zero error, i.e., to the state $\mathbf{x}=0$, in a minimum number of sampling periods.

If $\mathbf{u}_1$, $\mathbf{u}_2$ are the unit vectors along the co-ordinate axes in the state space of the running example, then the state vector at any time kT can be written in the form (see Fig. 4)

$$\mathbf{x}(k)=x_1(k)\mathbf{u}_1+x_2(k)\mathbf{u}_2$$

Now consider the question: From what initial states $\mathbf{x}(0)$, if any, is it possible to reach the origin of the state space in one sampling interval? Let $\mathbf{x}^1(0)$ be any initial state with this property; setting $\mathbf{x}(1)$ equal to zero in equation 4 leads to

$$\mathbf{x}(1)=0=G(T)\mathbf{x}^1(0)+m(0)\mathbf{h}(T)$$

Solving for $\mathbf{x}^1(0)$ and using equation 12 shows that

$$\mathbf{x}^1(0)=-m(0)G^{-1}(T)\mathbf{h}(T)=-m(0)G(-T)\mathbf{h}(T) \quad (13)$$

Letting

$$\mathbf{v}_1=G(-T)\mathbf{h}(T) \quad (14)$$

equation 13 shows that every state $\mathbf{x}^1(0)$ can be written in the form

$$\mathbf{x}^1(0)=y_1\mathbf{v}_1 \quad (15)$$

where y_1 is some real constant. The set of all states satisfying equation 15 clearly constitute a straight line in the state space.

The set of all states from which the origin can be reached in at most two steps is clearly the same set from which a state of the form given by equation 15 can be reached in at most one step. Let $\mathbf{x}^2(0)$ be a state from which the origin can be reached in two steps; then, again using equation 4,

$$\mathbf{x}(1)=y_1\mathbf{v}_1=G(T)\mathbf{x}^2(0)+m(0)\mathbf{h}(T)$$

Solving for $\mathbf{x}^2(0)$ and again using equation 12 as well as equation 14 leads to

$$\mathbf{x}^2(0)=G^{-1}(T)(y_1\mathbf{v}_1-m(0)\mathbf{h}(T))=y_1G(-2T)\mathbf{h}(T)-m(0)G(-T)\mathbf{h}(T)$$

Letting $\mathbf{v}_2=G(-2T)\mathbf{h}(T)$, it follows from equation 14 that every $\mathbf{x}^2(0)$ can be written in the form

$$\mathbf{x}^2(0)=y_1\mathbf{v}_1+y_2\mathbf{v}_2 \quad (16)$$

where y_1, y_2 are real numbers. Using the

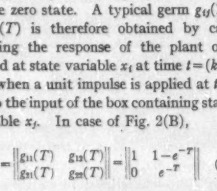


Fig. 3. Trajectory in state space

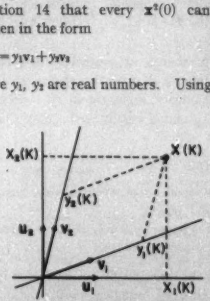


Fig. 4. Co-ordinate systems

explicit forms of $G(T)$ and $\mathbf{h}(T)$ given by equations 7 and 8, $\mathbf{v}_1$, $\mathbf{v}_2$ can be written as linear combinations of the unit vectors $\mathbf{u}_1$, $\mathbf{u}_2$

$$\mathbf{v}_1=(T+1-e^T)\mathbf{u}_1+(e^T-1)\mathbf{u}_2$$
$$\mathbf{v}_2=(T-e^{2T}+e^T)\mathbf{u}_1+(e^{2T}-e^T)\mathbf{u}_2 \quad (17)$$

It is easily verified from equation 17 that the vectors $\mathbf{v}_1$, $\mathbf{v}_2$ are linearly independent for any value of T, i.e., they do not lie along the same line through the origin of the state space. Therefore $\mathbf{v}_1$, $\mathbf{v}_2$ may be regarded as unit vectors along the axes of a new co-ordinate system, $\mathbf{x}(0)$ having the co-ordinates y_1, y_2 in this new system, as given by equation 16. Since any vector in the two-dimensional state space of the running example can be represented as the linear combination of at most two linearly independent vectors, it follows also from equation 16 that the origin can be reached always in at most two steps.

In the n^{th} order case, it can be rigorously shown (in the previously mentioned unpublished paper) that the set of n vectors

$$\mathbf{v}_1,\ldots,\mathbf{v}_n \text{ where } \mathbf{v}_i=G(-iT)\mathbf{h}(T) \quad (18)$$

are linearly independent if and only if

$$T\omega\neq q\pi \quad (19)$$

where q is any integer and ω is the imaginary part of any of the complex poles of the transfer function of the plant. The co-ordinate system for which the $\mathbf{v}_i$ in equation 18 are the unit vectors will be called the canonical co-ordinate system.

The foregoing may be summarized as follows:

If the sampling period T satisfied condition 19, then any state $\mathbf{x}$ may be written in the form

$$\mathbf{x}=\sum_{i=1}^{n}y_i\mathbf{v}_i \quad (20)$$

where $\mathbf{v}_i=G(-iT)\mathbf{h}(T)$ and n is the order of the system. Moreover, if j is the largest integer for which $y_j\neq 0$, then the origin may be reached from $\mathbf{x}$ in exactly j steps.

It should be noted also that the maximum response time of an n^{th} order optimal system is nT seconds. Thus the response time of a sampled system can be made arbitrarily small by selecting a short sampling period. However, by doing so, the magnitude of the control signals required will increase rapidly and the fast response may not be obtainable because plant saturation will ultimately limit the effect of the control signal.[6] As a rough rule, it probably does not pay to decrease T below about 10% of the dominant time constants of the plant.

The Control Signal Sequence

The next problem is to determine the actual values of $m(0)$, ..., $m(n)$ which will take an arbitrary initial state $\mathbf{x}(0)$ into the origin. Considering again the running example, it is known that every initial state can be written in the form:

$$\mathbf{x}(0)=y_1(0)\mathbf{v}_1+y_2(0)\mathbf{v}_2 \quad (21)$$

If $m(0)$ is chosen in an optimal way, the state at the next sampling instant must be given by

$$\mathbf{x}(1)=y_1(1)\mathbf{v}_1 \quad (22)$$

Equation 4 for $k=0$ is

$$\mathbf{x}(1)=G(T)\mathbf{x}(0)+m(0)\mathbf{h}(T)$$

Substituting equation 21 for $\mathbf{x}(0)$

$$=G(T)(y_1(0)\mathbf{v}_1+y_2(0)\mathbf{v}_2)+m(0)\mathbf{h}(T)$$

Using equation 18, the definition of the $\mathbf{v}_i$, and also equation 11, it further follows

$$\mathbf{x}(1)=y_2(0)\mathbf{v}_1+(y_1(0)+m(0))\mathbf{h}(T)$$

Hence $\mathbf{x}(1)$ will have the form required by equation 22 if and only if

$$m(0)=-y_1(0) \quad (23)$$

in which case

$$y_1(1)=y_2(0) \quad (24)$$

If $\mathbf{x}(1)$ is to be reduced to zero in one step

$$\mathbf{x}(2)=0=y_1(1)G(T)\mathbf{v}_1+m(1)\mathbf{h}(T)$$
$$=(y_1(1)+m(1))\mathbf{h}(T)$$

Therefore

$$m(1)=-y_1(1)=-y_2(0) \quad (25)$$

These results can be easily generalized as follows:

At every sampling instant, the optimal control signal $m(k)$ is given by:

$$m(k)=-y_1(k) \quad (26)$$

where

$$\mathbf{x}(k)=\sum_{i=1}^{n}y_i(k)\mathbf{v}_i \quad (20)$$

The sequence of control signals $m(k)$ which takes an arbitrary initial state $\mathbf{x}(0)$ to the origin in the absence of any disturbances is given by

$$m(0)=-y_1(0),\ m(1)=-y_2(0),\ \ldots,$$
$$m(n-1)=-y_n(0) \quad (27)$$

If $m(k)$ satisfies equation 26, then

$$\mathbf{x}(k+1)=\sum_{i=1}^{n-1}y_{i+1}(k)\mathbf{v}_i \quad (28)$$

and therefore

$$y_i(k+1)=y_{i+1}(k) \quad (29)$$

A simple and interesting geometrical interpretation of equations 28 and 29 is this. The initial state of the plant may lie anywhere in the n-dimensional state space. After one sampling period and the application of a control signal as given by equation 26, the state can be expressed as a linear combination of only $n-1$ linearly independent vectors $\mathbf{v}_1$, ..., $\mathbf{v}_{n-1}$; in other words, after one sampling period, the initial state always lies in an $(n-1)$-dimensional space, provided there are no external disturbances. After two sampling periods, the initial state is taken into a space of one less dimension during each sampling period. Ultimately, the state must be in a space of dimension zero which is an equilibrium point in the state space.

Physical Realization of the Control Signal

According to equation 26, the optimal control signal is obtained by computing the first component in the canonical co-ordinate system of the state of the plant at each sampling instant. In general, the state variables cannot be directly measured in the canonical co-ordinate system; on the other hand, any component of a vector in one co-ordinate system can be expressed as a linear combination of the components of the same vector in a different co-ordinate system. Thus it is always possible to write:

$$m(t)=-y_1(t)=-\alpha_i x_i(t)=\ldots-\alpha_n x_n(t)$$
$$\text{(all } t\text{)} \quad (30)$$

where the α_i are real constants and the x_i are directly measurable state variables. To calculate these constants, observe that

$$\sum_{i=1}^{n}x_i\mathbf{u}_i=\sum_{i=1}^{n}y_i\mathbf{v}_i \quad (31)$$

by definition. Moreover, since $\mathbf{v}_j=G(-jT)\mathbf{h}(T)$

$$\mathbf{v}_j=\sum_{i=1}^{n}p_{ij}\mathbf{u}_i$$

where

$$p_{ij}=\sum_{l=1}^{n}g_{il}(-jT)h_l(T) \quad (32)$$

The $g_{il}(-jT)$ are the elements of the transition matrix in the physical co-ordinate system as given by equation 6. Likewise, the $h_l(T)$ are components of $\mathbf{h}(T)$ in the physical co-ordinate system as given by equation 8. Substituting into equation 31

$$\sum_{i=1}^{n}y_i\mathbf{v}_i=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}p_{ij}y_j\right)\mathbf{u}_i=\sum_{i=1}^{n}x_i\mathbf{u}_i \quad (33)$$

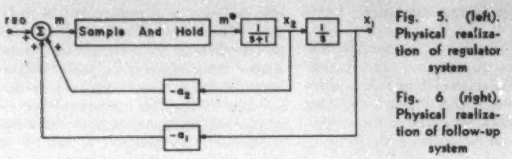
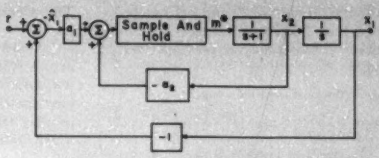


Fig. 5. (left). Physical realization of regulator system

Fig. 6 (right). Physical realization of follow-up system



Since the $\mathbf{u}_i$ are linearly independent, equation 33 implies

$$\sum_{j=1}^{n} p_{ij}y_j = x_i \text{ (all } i\text{)} \tag{34}$$

Thus the α_i can be obtained by solving the set of linear equations 34 for y_1. (Because the $\mathbf{v}_i$ were chosen so as to form a co-ordinate system, equation 34 always has a unique solution.)

In the running example, see equation 17,

$$P = \begin{Vmatrix} T+1-e^T & T-e^{2T}+e^T \\ e^T-1 & e^{2T}-e^T \end{Vmatrix}$$

Therefore

$$y_1 = \frac{p_{22}}{\det P}x_1 + \frac{(-p_{12})}{\det P}x_2$$

Comparing with equation 30, the final result is seen to be

$$\alpha = e^T/T(e^T-1),$$
$$\alpha_2 = (e^{2T}-e^T-T)/T(e^T-1)^2$$

Fig. 5 shows the block diagram of the optimal regulator system when $r(t)=0$.

Nonzero Inputs

There are several interesting problems which must be discussed if the solution for the optimal regulator system is to be extended to follow-up systems.

Arbitrary Step Inputs

This problem is equivalent to making the output c equal to some arbitrary reference input r in such a way that the plant is in equilibrium with zero system error.

In the regulator problem previously treated, this equilibrium point with zero error is the origin of the state space.

Provided the state x_1 is the system output, zero error at equilibrium is achieved when $x_1=r$ and the other components of the state vector are determined by the equilibrium conditions.

If, as in the running example, the plant contains at least one integration, then the differential equations of the plant in terms of the state variables may be written as follows:

$$dx_1/dt = x_2 \tag{35}$$

plus other equations which do not involve the state variable x_1. Form the modified state variables

$$\hat{x}_1 = x_1 - r$$
$$\hat{x}_k = x_k \qquad 1<k\leq n \tag{36}$$

Since $dr/dt=0$, substituting equation 36 into equation 35 leads to an equation of the same form as equation 35 in $\hat{x}_1$. The rest of the equations of the plant are unaffected. In equilibrium,

$$\hat{x}_1 = r$$
$$\hat{x}_k = 0 \qquad 1<k\leq n$$

which shows that the follow-up action is achieved with the regulator design by utilizing the modified state x_1-r in place of x_1.

Thus the regulator system of Fig. 5 which contains an integration is converted to a follow-up system simply by taking the difference between state x_1 and the input r to create the modified state $\hat{x}_1$ prior to multiplying by the feedback coefficient α_1. This system is shown in Fig. 6.

In those cases where the plant does not contain an integration proceed as follows. From Laplace transform considerations note that in the steady state the control signal is given by

$$m_{ss} = r/g(0) \tag{37}$$

where $g(s)$ is the transfer function of the plant. Using equation 4, it is seen that the equilibrium state $\mathbf{z}$ is given by

$$\mathbf{z} = G(T)\mathbf{z} + m_{ss}\mathbf{h}(T) \tag{38}$$

By a shift in the origin of the canonic co-ordinate system, it follows that any state may be written in the form

$$\mathbf{x}(k) = G(-nT)\mathbf{z} + \sum_{i=1}^{n} y_i(k)\mathbf{v}_i \tag{39}$$

where the $\gamma_i(k)$ are unique real constants. If, as before, the control signal is given by

$$m(k) = -y_1(k) \tag{26}$$

$y_1(k)$ to be determined by equation 39, then after one sampling period

$$\mathbf{x}(k+1) = G(-(n-1)T)\mathbf{z} + \sum_{i=1}^{n} y_i(k+1)\mathbf{v}_i \tag{40}$$

where $y_i(k+1) = y_{i+1}(k)$. Multiplying both sides of equation 38 by $G(-nT)$ and rearranging shows

$$G(-nT)\mathbf{z} = G(-(n-1)T)\mathbf{z} + m_{ss}\mathbf{v}_n$$

Substituting this into equation 40 results in

$$\mathbf{x}(k+1) = G(-nT)\mathbf{z} + \sum_{i=1}^{n} y_i(k+1)\mathbf{v}_i$$

where

$$y_i(k+1) = y_{i+1}(k) \quad i=1, \ldots, n-1 \tag{41}$$
$$y_n(k+1) = -m_{ss}$$

This shows that expression 39 can be used to compute $y_1(k)$ for every value of k. The foregoing may be summarized as follows.

For step inputs of arbitrary magnitude, the optimal control function is given by

$$m(k) = -y_1(k)$$
$$= \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} g_{ij}(-nT)z_j - x_i \right) \tag{42}$$

$\mathbf{z}$ being determined by equation 37, and the α_i by equation 34.

Moreover, if j is the largest integer in equation 39, such that $y_j \neq -m_{ss}$, then equilibrium is reached in exactly j steps. This shows that the modified state variables are given in general by $\hat{\mathbf{x}} = \mathbf{x} - G(-nT)\mathbf{z}$.

If the plant does not contain an integration, then its transition matrix has no characteristic root equal to one. Consequently the matrix $1-G(T)$ can be inverted and $\mathbf{z}$ can be found from

$$\mathbf{z} = m_{ss}(1-G(T))^{-1}\mathbf{h}(T) \tag{43}$$

In this case, however, the accuracy of the output in the steady state depends on the accuracy with which the transition matrix is known. Such systems are called calibrated and are frequently undesirable in practice.

If the plant does contain an integration, then $m_{ss}=0$. In this case, the transition matrix has a characteristic root equal to one and therefore there is a vector $\mathbf{z}$ satisfying

$$G(T)\mathbf{z} = \mathbf{z} \text{ and hence also } G(-nT)\mathbf{z} = \mathbf{z} \tag{44}$$

By substituting these relations into equation 42, it follows that the modified state

variables are now given by $\hat{\mathbf{x}}=\mathbf{x}-\mathbf{z}$. Since it is desired to have $x_1=c=r$ in the steady state, when $\hat{\mathbf{x}}=0$, $z_1=r$ and the rest of the components of the vector $\mathbf{z}$ are then determined uniquely by equation 44. Hence if there is an integration in the plant, the steady-state accuracy does not depend on precise knowledge of the transition matrix.

Arbitrary Polynomial Inputs

Suppose the permissible types of inputs are given by equation 3. Assume, for simplicity, that the system contains at least m integrators. Then the differential equations of the plant in terms of the state variables are

$$\begin{aligned} dx_1/dt &= x_2 \\ dx_2/dt &= x_3 \\ &\cdots \\ dx_m/dt &= x_{m+1} \end{aligned} \tag{45}$$

plus other equations which do not involve the state variables $x_1, \ldots, x_m$. If the modified state variables are defined as

$$\begin{aligned} \hat{x}_1 &= x_1 - r \\ \hat{x}_2 &= x_2 - dr/dt \\ &\cdots \\ \hat{x}_m &= x_m - d^m r/dt^m \\ \hat{x}_k &= x_k \qquad m<k\leq n \end{aligned} \tag{46}$$

then substituting into equation 45 shows that the modified state variables obey the same differential equations. Hence in equilibrium when $\hat{\mathbf{x}}=0$,

$$\begin{aligned} x_1 &= r \\ x_2 &= dr/dt \\ &\cdots \\ x_m &= d^m r/dt \end{aligned}$$

which shows that the output will be identical with the input after, at most, n sampling periods. Note, however, that equations 46 require the knowledge of the derivatives of the input. This may be difficult to obtain in some cases.

Some Practical Aspects

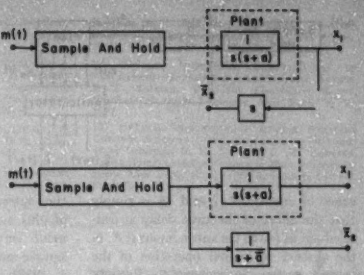
Simulation of the State Variables

In some practical cases it is not possible to measure directly all of the state variables. With such plants other elements must be added in order to obtain approximations, $\hat{x}_i$, to those state variables which are inaccessible. Two common methods of obtaining the approximate state variables are the following.

1. If $g_{ij}(s)$ is the transfer function relating state variable x_i, which is measurable, to state variable x_j, which is not measurable, then $\hat{x}_j$ may be obtained by passing the measurable signal x_i through the transfer function $1/g_{ij}(s)$.

2. Construct a model or analog of the plant in which all of the desired state variables can be measured, and subject it to the same input sequence as that applied to the actual plant. Because the output of the plant model is required only at sampling instants, the model can be simulated also by a digital computer.

Fig. 7. Methods for simulating state variables



Both of these methods have been quite well known in the control systems art, although their full significance was perhaps not understood in the past. In fact, method 1 includes the very popular tachometer-feedback technique of stabilization; see also the experimental results which follow. The use of models to simulate the behavior of the plant has also received much attention in recent years; see, for instance, a recent paper by Reswick.[8] Finally, common methods of compensation of sampled-data systems may also be regarded as special cases of method 2.

To illustrate the application of these methods, assume that in the running example it is not possible to directly measure the state variable x_2. Fig. 7 shows how the variable $\hat{x}_2$ may be obtained by either method 1 or 2. Using method 1, the state variable x_1, which can be measured, is differentiated to obtain an approximation to x_2 which, by equation 5, is dx_1/dt. Using method 2, the sampled control signal $m^*(t)$ is applied to the element whose Laplace transform is $1/(s+a)$.

In method 1 the transfer functions required to obtain the approximations have at least a simple pole at ∞. In practice, such elements are not realizable although tachometers and accelerometers are reasonable approximations over a certain range of frequencies. Because of this difficulty of physical realization, method 2 is frequently more satisfactory in practice.

In using method 2, note that $\hat{x}_2=x_2$ only if the model is identical with the plant element it is simulating ($a=\bar{a}$) and only if there are no (disturbance) inputs applied to the plant which do not also affect the model in the same way. Clearly, the best results are achieved by measuring as many state variables as possible. In some practical cases, it may be possible to ignore some of the state variables, since they may correspond to time constants very much smaller than T and therefore differ negligibly from their steady-state values at the sampling instants. A detailed study of the approximation problem will be made at a later time.

Compensation of Plant Time Delay

Suppose the control signal affects the plant only after a definite time delay τ. This may happen particularly often in process control problems where the time delays are caused by transportation of liquids and gases through long pipes. As is well known, such time delays may cause serious problems in the stabilization of the closed-loop system. A simple way of canceling the detrimental effect of time delay as far as over-all system stability is concerned is the following.

Select the sampling period T so as to have $\tau=qT$, where q is a positive integer. Since the control signal affects the plant only after qT seconds, it should be computed not according to the present state of the plant but according to the anticipated state of the plant qT seconds later. Thus instead of equation 26 the control signal should be given by:

$$m(k) = -y_1(k+q) \tag{47}$$

where $y_1(k+q)$ is the first component of the approximate state vector $\mathbf{x}(k+q)$ in the canonical co-ordinate system.

$$\bar{\mathbf{x}}(k+q) = G(qT)\mathbf{x}(k) + \sum_{i=1}^{q} m(k-i)G((i-1)T)\mathbf{h}(T) \tag{48}$$

Equations 48 characterize the anticipator in Fig. 8, which gives the value of the state q sampling periods after the present sampling instant, provided that no disturbances act on the plant during the interval $kT\leq t\leq(k+q)T$. The signifi-

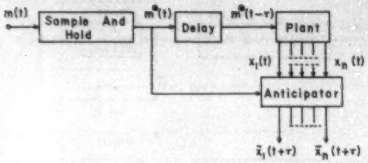


Fig. 8. Compensation for delay

cance of equations 47 and 48 is simply that the effect of the time delay is canceled out as far as the autonomous (i.e., in the absence of inputs) operation of the control system is concerned. Because the plant is assumed linear, the effect of disturbances during the anticipation interval is superposed on the autonomous behavior. If these disturbances are completely random, that is, completely unpredictable, then a system containing a plant with time delay clearly cannot have optimally fast disturbance response. If the disturbances can be predicted in some statistical sense, then the predictor equations for the disturbance are to be incorporated in the equations of the anticipator, thereby improving the disturbance response of the system.

The anticipator is physically realizable by either analog or digital computation since equation 48 requires only knowledge of the present state, a sequence of q past control signals, and knowledge of the elements of the transition matrix. The latter requirement shows that this compensation scheme is of the "calibrated" type. However, as long as the plant contains an integrator, the output will be fed back with coefficient -1, so that the steady-state accuracy does not depend on the precise knowledge of the transition matrix.

This method of compensation of the effects of time delay was first suggested in a general form by Bass.[9]

Experimental Results

An instrument servomechanism shown in Fig. 9 was constructed to perform in the optimal fashion. The components of this system are conventional instrument servo hardware except for the sample-and-hold element which employs a slight modification of the 6-diode gate of reference 10. The error signal was d-c and was sampled and held at a rate of 10 cycles per second. This signal was then used to modulate a 60-cycle-per-second carrier in order to drive a 2-phase servomotor. As is well known, the signal-frequency transfer function of such motors is approximately of the same form as that of the running example. One state variable was the output; an approximation to the other was obtained by attaching a tachometer to the output shaft of the motor.

The adjustment of K_1 and K_2 was found to be extremely simple. In fact, optimal settings could be made in a few minutes by persons unfamiliar with the operation of the system, once the criterion of "deadbeat response" was understood. The calculated and experimentally derived settings for optimal performance were found to differ by only a few per cent. The discrepancy is due mainly to the neglect of additional state variables in the system. See Table I.

Table I

	K_1	K_2
Calculated (based on measurement of motor transfer function)	14.5	0.0712
Experimental	13.9	0.074

Conclusions

This paper presents a method for designing a sampled-data control system having the following properties:

1. The design procedure is applicable to any linear plant with a rational transfer function.

2. The response of the system to any disturbance in state or step input requires at most n sampling periods, n being the order of the plant.

3. After at most n sampling periods following the application of a disturbance or step, the system reaches equilibrium with zero error. Hence the system is not subject to intersample ripple.

4. The design of the system is based directly on the concept of the state and can be readily implemented if all states are accessible.

5. If not all states are accessible the design should incorporate some method of simulating the state which usually requires analog or digital computation.

6. Additional computation is necessary in order to compensate the effect of time delay.

7. The design results in a linear system. It can be shown, however, that even if the control signal is limited by saturation, i.e.,

$$|m^*(t)| \leq M$$

the system is still monostable.[11] In other words, any initial state is reduced to equilibrium (in a finite number of steps) if the system is designed on a linear basis.

8. The optimal design may be extended to include the case of saturation. In this case it can be shown[12] that the optimal control signal $m(k)$ is in general a nonlinear function of the state vector $\mathbf{x}(kT)$.

It is hoped that the present discussion will help in simplifying the design of computer-control systems and lead to a better understanding of the reasons why real-time computation can contribute to better control. Using many other types of criteria for optimality, one can arrive at other optimal designs such that the control signal is a linear function of the state.[13] Reference 13 also gives an alternate method for the derivation of the feedback constants α_i in equation 30.

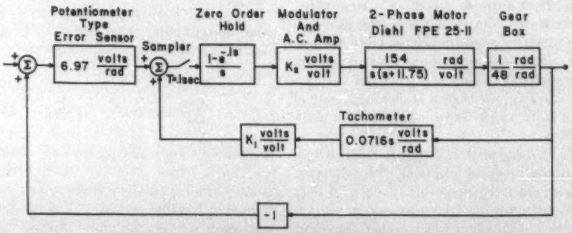


Fig. 9. Block diagram of practical system

References

1. The Pulse Transfer Function and Its Application to Sampling Servo Systems, R. H. Barker. *Proceedings*, Institution of Electrical Engineers, London, England, *Monograph no. 43*, pt. 4, 1952.

2. Sampled-Data Processing Techniques for Feedback Control Systems, A. R. Bergen, J. R. Ragazzini. *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 336–47.

3. Factors in the Design of Digital Controllers for Sampled-Data Feedback Control Systems, J. E. Bertram. *Ibid.*, vol. 75, pt. II, July 1956, pp. 151–59.

4. Compensation of an Error-Sampled System

by a Multirate Controller, G. M. Krane. *Ibid.*, vol. 76, pt. II, July 1957, pp. 149–59.

5. Synthesis and Critical Study of Sampled-Data Control Systems, E. I. Jury. *Ibid.*, vol. 75, pt. II, July 1956, pp. 141–51.

6. *Discussion* by R. E. Kalman of reference 2, pp. 245–46.

7. Theory of Ordinary Differential Equations (book), E. A. Coddington, N. Levinson. McGraw-Hill Book Company, Inc., New York, N. Y., 1953, chap. 3.

8. Disturbance Response Feedback—A New Concept, J. B. Roswick. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 78, 1956, pp. 153–62.

9. A Generalization of the Functional Relation $Y(t+h) = Y(t) \cdot V(a)$ to Piecewise-Linear Difference-Differential Equations, W. R. Bass. *Quarterly of Applied Mathematics*, Providence, R. I., vol. 14, 1957, pp. 415–17.

10. Pulse and Digital Circuits (book), J. Millman, H. Taub. McGraw-Hill Book Company, Inc., New York, N. Y., 1956, p. 445.

11. Physical and Mathematical Mechanisms of Instability in Nonlinear Automatic Control Systems, R. E. Kalman. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 79, 1957, pp. 553–56.

12. Optimal Nonlinear Compensation of Saturating Systems by Intermittent Action, R. E. Kalman. *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., 1957.

13. Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes, R. E. Kalman, R. W. Koepcke. *Transactions*, American Society of Mechanical Engineers, New York, N. Y. (To be published in vol. 81, 1959).

Trends in Automatic Hoist Controls

G. L. TILEY
ASSOCIATE MEMBER AIEE

E. ZUCKER
ASSOCIATE MEMBER AIEE

Synopsis: The paper outlines the problems facing mine hoist engineers in the design and application of automatic hoist equipment. Duty cycle requirements, types of hoists, and the advantages of mine hoist automation are discussed. Particular mention is made of available a-c and d-c control systems. Overriding speed reference equipment and the safety features of modern automatic controls are described. A discussion on hoist brakes is also included, as these are an integral part of each hoist control design. It is the purpose of this paper, and its bibliography, to provide the designer with many references and with a detailed outline of existing thinking on the subject.

MOST of the mineral requirements of North America have been obtained, up to the present, from ore bodies that can be worked by open cast methods. Supplies of readily available ore are running out, however, and it is becoming necessary to go deeper into the earth for the required materials. As a result, increasing attention is being paid to the design of mine hoists, their drives, and associated control systems.

Control Requirements

A thorough investigation of the hoisting requirements of any particular installation should be made before hoisting or control equipment is selected. References 1 to 5 should be consulted on the general subject of mine hoists and mine-hoist duty cycles.

The basic requirements of the hoisting cycle from the control point of view are simple. The hoist must be accelerated, run at full speed, decelerated, and finally stopped. On automatic control, the difficulties arise chiefly in the required degree of accuracy with which the hoist conveyances must be stopped. A multilevel cage hoist is often required to have a landing accuracy of about one inch in a 4,000-foot shaft under all conditions of speed and loading. The difficulty of accomplishing this is further aggravated by the fact that the connection between the hoist and the conveyances is by relatively elastic ropes which are also subject to permanent deformation (rope stretch) creeping around the circumference of a Koepe wheel, etc.

The controls must also fail to safety under all imaginable circumstances. The effectiveness with which safety requirements have been met on present-day mine hoists is illustrated by the fact that these have a better safety record than any other known form of transportation.

Types of Hoists

Figs. 1(A), 1(B), and 1(C) show the three most common types of hoists. The conventional drum hoist [Fig. 1(A)] is admirably suited for service in shafts less than 2,500 feet deep. Two shaft conveyances are attached to ropes that coil on separate drums so arranged that when one conveyance is being raised, the other is being lowered.

At depths of over 2,500 feet, the Koepe or friction hoists become attractive. These hoists, shown in Figs. 1(B) and 1(C), are relatively new to North America. The head rope or ropes are not attached to the hoist wheel but are only in contact with it over an arc of 180 to 210 degrees. Forces are transmitted from the wheel to the rope by friction only. Provided the ratio of rope tensions on either side of the wheel is carefully controlled, there is no danger of the rope slipping. A tail rope is connected between the bottoms of the conveyances in order to keep these tension ratios within safe limits. As a result, the only out-of-balance is the useful load being raised. On a conventional drum hoist, the unbalanced rope in the shaft imposes additional torques on the hoist motor.

When large tonnages have to be handled, the size of the single head rope required by either a drum hoist or a ground-mounted Koepe hoist becomes impractical. A Koepe hoist can, however, be mounted in a tower directly over the shaft, as shown in Fig. 1(C) and then two or four head ropes can be used. This system is similar to a multirope elevator.

As has already been noted, the unbalanced rope weight imposes additional loads on the conventional drum hoist drive. The inertia of the Koepe hoist is also considerably less than that of a double drum hoist, so the inertia loads during acceleration and deceleration will also be reduced.

Duty cycles of a Koepe and a drum hoist are compared in Fig. 2. Both hoists are designed for hoisting 400 tons per hour from a depth of 3,500 feet, using 13-ton-capacity skips. The drum hoist requires a single 2-inch-diameter rope and 14-foot-diameter drums. The friction hoist employs four ropes, each of $1^{1}/_{8}$-inch diameter, and has a drum diameter of only 10 feet 6 inches. The former hoist requires a driving motor of 3,500 horsepower (6,500-horsepower peaks) while the latter only needs 2,500 horsepower (3,500-horsepower peaks).

It is obvious that the Koepe hoist is very attractive at these depths. From the control engineer's point of view, however, the system introduces additional complexities. Torques, in particular braking torques, must be carefully controlled and periodical recalibration of the control equipment is necessary to compensate for rope creep.

Paper 58-1210, recommended by the AIEE Mining Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 26, 1958; made available for printing August 21, 1958.

G. L. Tiley and E. Zucker are with the Canadian Westinghouse Company, Ltd., Hamilton, Ont., Canada.

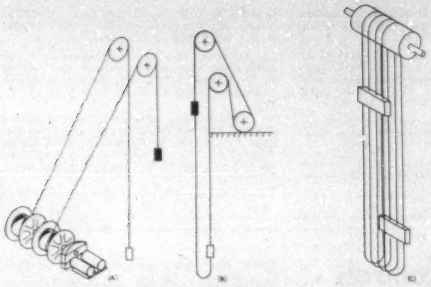


Fig. 1 (above). Common types of hoists

A—Conventional drum hoist
B—Ground-mounted Koepe hoist
C—Four-rope Koepe hoist over shaft

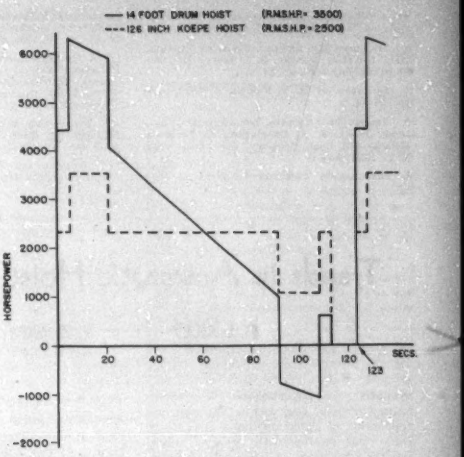


Fig. 2 (right). Comparison of duty-cycle drum and Koepe hoist

The introduction of the multirope tower-mounted Koepe hoist has led to a re-evaluation of the relative importance of the components supplied by the electrical and mechanical manufacturer. On a conventional drum hoist, driven by geared high-speed motors, the mechanical parts comprise 60–75% of the total equipment. The situation is reversed, however, when an automatic tower-mounted friction hoist is considered. When tower costs are included, a direct-drive hoist is often cheaper than a geared hoist. A direct-drive hoist may be supported in two bearings, with the motor armatures overhung, and a simple, relatively flexible tower construction employed. Such a hoist reduces to an inexpensive wheel, driven by expensive low-speed d-c motors (see Fig. 3), and represents the ultimate in reliability.

Automatic Control Considerations

The increased capital costs involved in an automatic hoist drive, and the fact that mining regulations usually require a skilled person to be in attendance, have deterred many mining operators from installing this type of equipment. Progressive companies are realizing, however, that they cannot afford not to install automatic or semiautomatic hoists. The case for automation can be summed up as follows.

Hoisting Efficiency

The mine shaft and its hoisting facilities represent a considerable capital investment. It is therefore imperative that the hoist perform as efficiently as possible. When the hoist is controlled by a human operator on surface, errors of judgment, natural caution, and fatigue all cause loss of production. On automatic control, the optimum cycle is achieved every time and tests have proved that production can be increased by as much as 15%.

The first step towards more efficient hoist control was taken on d-c manual hoists, a long time ago, by introducing automatic acceleration and slowdown features. Here the operator moves his master control lever to the "full speed" position and the hoist automatically accelerates, runs at full speed, decelerates at the proper point in the shaft, and comes to creep speed. All the operator does is apply the brakes and remove power from the hoist. It is thus logical to remove him altogether by controlling the skip through pushbuttons from the loading pocket.

With most types of ore a skip tender is necessary to supervise the loading arrangements, because the ore is not of uniform size and blockages are possible. Hoisting efficiency can be further increased by the use of underground crushers.

Cage (service) hoist operation also becomes more efficient once pushbutton control, elevator style, is introduced. The cage tender controls all movements from the cage, and signalling to and from surface is eliminated.

Manpower Saving

On most present automatic installations, a qualified hoist man is still retained to take over when the automatic controls are shut down. Shaft and rope inspection runs are done on manual control. A manpower saving is possible when two or more automatic hoists are located in the same hoist room. One man then supervises the multiple installation on surface. By definition, semi-automatic operation requires a man at the loading station to initiate each hoisting cycle. Full automatic hoisting means the hoist runs continuously without any supervision. The latter system, not in common use yet, may eventually save manpower at the underground location when materials to be hoisted are made suitable for this type of operation.

Safety and Equipment Protection

Elimination of the human element in signalling and control operations results in safer hoisting. Irregular or erratic control of the hoist is inevitable if inexperienced operators handle the equipment. Automatic controls, therefore, not only make hoisting safer but protect the hoist from abuse by poor manual operators.

Types of Hoist Control

D-C Ward-Leonard Control Using a Motor-Generator Set[3,6,7]

The variable-voltage d-c or Ward-Leonard system of control is ideal for mine hoists. Many successful systems of closed-loop control have been developed. High-gain static and rotating amplifiers

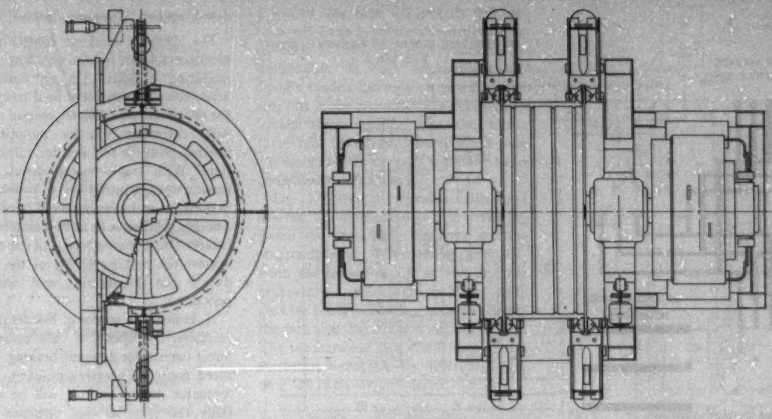


Fig. 3. Direct-drive friction hoist with unit disk brakes

are employed to give accurate control of hoist speed under all conditions of loading.

Designing a satisfactory speed regulator system for a mine hoist presents no particular difficulties. In most installations, only one motor has to be controlled. Sudden changes of speed are neither necessary nor desirable, so large forcing factors are not required.

The main criteria to be observed in the design of such a regulator can be summed up as follows:

1. Speed, particularly creep speed, must be accurately controlled under all conditions of loading.

2. Suiciding must be effective to prevent excessive circulating currents at standstill.

3. There must be overriding current limit control to protect the electric equipment against excessive currents, and the mechanical equipment against excessive torques. The current limit response should be virtually independent of pattern and voltage signals. This is particularly important when the hoist is being maneuvered under manual control. The operator must then have full torque available at all positions of the master controller.

The basic components of a typical Ward-Leonard system of controlling a d-c hoist motor are shown in Fig. 4. A speed reference signal determines the desired speed at any point in the shaft. This signal is amplified by a magnetic amplifier, and fed via directional contactors to the pattern field of a Rototrol exciter. The latter controls the field of the main generator which, in turn, controls the voltage of the armature loop. A signal, proportional to armature loop voltage, is fed back to the Rototrol voltage field and there matched against the pattern field. The difference between the two fields provides the net excitation to the Rototrol. The higher the gain of this exciter, the smaller the error under steady-state conditions. The hoist motor operates with a constant field. As a first approximation, it can be said that the motor speed is proportional to loop voltage.

A regulator as just described has a very high gain. Dangerously high currents could flow if the pattern signal were suddenly altered, because the inertia of the hoisting system would prevent the speed reaching steady-state conditions immediately. In other words, generator volts would tend to rise or fall much more rapidly than the back-electromotive force of the motor, and excessive currents would flow in the armature loop circuit.

In this system, the current-limiting device takes the form of two magnetic amplifiers in push-pull. They are excited by a signal proportional to the armature loop current, and biased to give no output until the prescribed current limit is reached. The output of these magnetic amplifiers acts to keep the net flux in the Rototrol fields down so that design values of current cannot be exceeded in the armature loop.

A further refinement is to introduce a signal to these current-limit magnetic amplifiers proportional to the loop voltage, once this exceeds a certain value. This

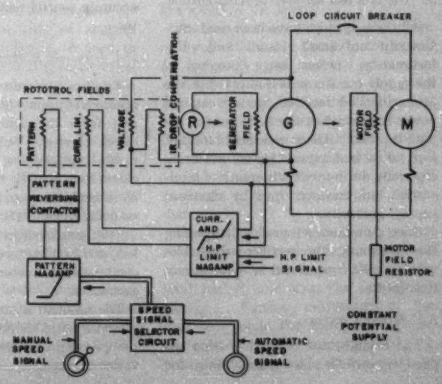


Fig. 4. Basic Ward-Leonard hoist control circuit

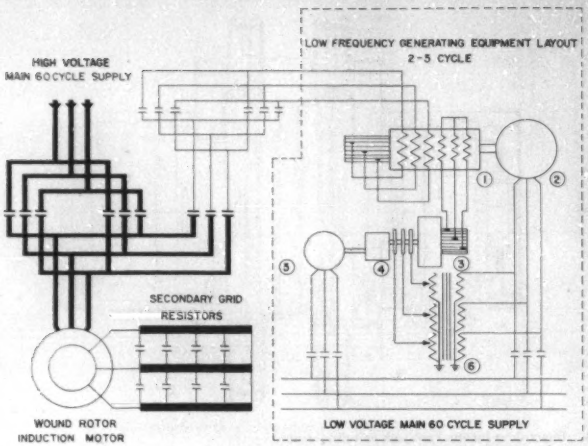


Fig. 5. Low-frequency dynamic braking control circuit

has the effect of "backing off" the current limit setting above a certain speed, and thus reduces the peak horsepower demands on the motor-generator set driving motor. It also insures a smooth transition from acceleration to full-speed run.

D-C Ward-Leonard Control Using a Grid-Controlled Rectifier[7,8]

Several schemes have been developed in which the motor-generator set in the Ward-Leonard system is replaced by a grid-controlled mercury-arc rectifier. Motor reversal may be effected by:

1. Reversal of the hoist motor field.
2. Reversal of the hoist-motor armature supply.
3. Use of a rect fier-inverter combination.

All these systems have been used satisfactorily on small installations. Unfortunately, unless tap changing of the supply transformers is resorted to, the grid-controlled rectifier operates at bad power factors when the hoist runs at reduced speeds. These periodical blocks of load at a bad power factor, occurring especially during the early stages of acceleration, are frowned upon by electrical supply authorities.

Once a method is found of overcoming this drawback, the rectifier adjustable-voltage hoist control will have an attractive future.

A-C Hoist Control

The wound-rotor motor has been applied successfully to a-c manual hoists for many years. This drive costs considerably less than the Ward-Leonard system. On duty cycles in which the hoist is running at full speed for the major portion of the cycle, operational costs may also be lower as losses occur on only one machine.

Direction of hoisting is controlled by stator reversing contactors, and speed by secondary resistance in the form of metallic resistor grids and contactors, or a liquid rheostat. The former is more common in North America. The main problems of wound-rotor motor control arise during slow-speed operation and electric braking. Counter torque braking (plugging) may be acceptable for manual hoists, but it is wasteful of power and accurate control requires a skilled operator.

D-C Dynamic Braking[6,7,9]

Braking performance can be greatly improved by removing the alternating current from the stator and substituting direct current. Very effective d-c dynamic braking schemes, embodying closed-loop control, have been developed by several European and North American manufacturers. The European schemes usually employ liquid secondary controllers, and braking torque is regulated by a combination of rotor resistance and d-c excitation.

The problem is rather more difficult if secondary grid resistors are used, because the resistance must, of necessity, be varied in steps.

Low-Frequency Dynamic Braking[7,10,11]

D-c dynamic braking control is an excellent solution to the problem of decelerating a hoist. On an automatic hoist, however, control of final creep into the dump becomes difficult, as load conditions determine whether motoring or braking power is required for this operation. With single motor drives, load sensing must be introduced to switch from braking to motoring power if necessary. Another solution is to fit two motors, and balance the motoring torque of one motor against the braking torque of the other during the deceleration and creeping period.

If, however, a source of low frequency (3 phase, 2–5 cycles) is used instead of direct current for dynamic braking, creep speed control is greatly simplified. Performance during braking will be essentially the same as that obtained with direct current. When the motor approaches the synchronous speed of the low-frequency supply (3–10% of full speed), the rotor is short-circuited and the low-frequency voltage is increased. The motor will then run at very near synchronous speed, regardless of whether the load is motoring or braking.

Fig. 5 shows the basic components of the low-frequency dynamic braking scheme developed by Canadian Westinghouse Company. The low-frequency supply is obtained from a low-frequency generator (*1*) that is excited by a frequency converter (*3*). Reversing contactors select the sense of rotation of the low-frequency supply in a similar manner to the primary reversing contactors. The system has been discussed in detail in recent AIEE papers.[10,11]

If a combination of resistance and reactance is used in the rotor circuit, the speed torque curves, for any particular value of low-frequency voltage, are essentially flat, as shown in Fig. 6. With such an arrangement, a relatively simple closed-loop scheme for automatic a-c hoists is possible, as braking effort may be regulated by control of the low-frequency supply voltage only.

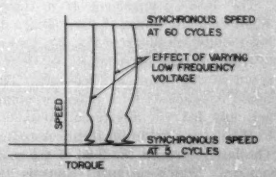


Fig. 6. Speed-torque curves on low frequency

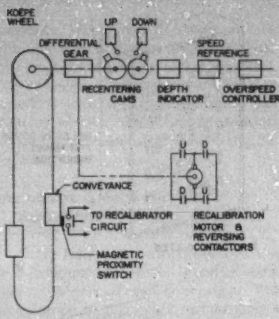


Fig. 7. Koepe hoist recalibration scheme

Primary Saturable Reactor Control[12,13]

Very sensitive control schemes have been developed for a-c powered cranes, bridges, and lock gates. These employ a wound-rotor induction motor with secondary resistors, and with saturable reactors in the primary circuit. During braking and slow-speed running, the rotor resistance is set to give, say, pull-out torque at standstill. At the same time, the stator voltages are either unbalanced, or reduced symmetrically, to give an infinite range of torque within the envelope of the full stator voltage characteristics.

This scheme has been applied to motors of up to 150 horsepower in North America, but the Japanese have a 500-horsepower hoist with primary saturable-reactor control operating at the Oyuban Mine of the Mitsubishi Mining Company.

The authors venture to predict that the a-c hoist of the future will be driven by a force-ventilated squirrel-cage induction motor, with saturable-reactor primary control.

Control of a Koepe Hoist

Control of a Koepe hoist is essentially similar to that of a drum hoist. Care must be taken to insure that excessive torques are not applied to the system, as otherwise there is a danger of the rope slipping. Some means of detecting rope slip should be provided. This usually takes the form of two speed-sensitive switches, one driven by the drum and one by a pulley that rides on the rope. Major speed differences between these switches cause an emergency stop.

There is a tendency for the head ropes to creep around the Koepe wheel, and throw the controls out of step with the conveyances. This phenomenon should not be confused with rope slip. It results from the fact that when a conveyance is loaded, the rope will stretch and the hoist will have to rotate an extra angular distance to bring the conveyance to its destination. If both conveyances are equally loaded the rope creep will cancel out, but if one conveyance is persistently loaded more heavily than the other, there will be a resultant creep in one direction and the controls will have to be periodically recalibrated.

Fig. 7 shows a typical recalibration scheme. A differential gear is fitted between the drive from the hoist and the control devices. At the end of each cycle, when the conveyance is at rest and the brakes are applied, limit switches operated by the recalibrating cams will detect the fact that the controls are out of step. The recalibrating motor will then feed a correction signal into the differential gear until the cams are recentered.

Speed Reference Systems[7,14]

Control for an Automatic Hoist

Some method of generating a speed reference or pattern signal is required for any closed-loop control scheme. The earliest speed-reference control equipment employed cam-operated limit switches, driven from the hoist drums. These are set to increase or decrease the speed in a number of steps. Accuracy may be improved by the use of traveling nut limit switches, or long-range cams. Such systems are excellent when hoisting from one level only, but become complicated when a number of levels have to be served.

An ingenious development by Allmänna Svenska Elektriska Aktiebolaget (ASEA) utilizes a speed controller based on the "hunting tooth" principle. Magnets are mounted on adjacent gear wheels running at slightly different speeds. After a given number of revolutions, the magnets coincide to operate a pair of magnet switches. The controller is designed to give four speed points during acceleration and deceleration at each level, and the whole assembly is compact and reasonably easy to service.

An entirely new approach to the problem of providing a suitable speed-reference signal for a multilevel automatic hoist is embodied in the Westinghouse advance selector. This equipment, based on high-speed passenger elevator practice, is, in effect, a differential gear that integrates both a speed and a position signal.

Fig. 8 shows the basic components of the advance selector for a Koepe hoist (duplex advance selectors for double-drum hoists have also been developed). An endless chain T connects sprocket wheels A, B, C, D, E, F, and G. Sprocket D is driven by the hoist shaft, and rotation of this sprocket will advance the conveyance carriages E and F by an amount proportional to shaft conveyance travel.

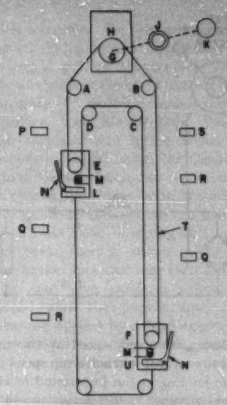


Fig. 8. The advance selector components

An infinitely variable speed-reference signal is generated by the variable inductor J, which is attached to sprocket G and is driven through a clutch by a small motor K. The inductor has a single-phase stator and a single-phase rotor winding, and the output voltage is proportional to the angular displacement of the rotor.

Rotation of sprocket G will advance the conveyance carriages by an amount proportional to the speed reference signal. The conveyance carriages, then, are advanced with respect to the shaft conveyances by an amount proportional to the speed of the hoist.

At the commencement of the cycle, motor K will rotate the inductor to call for increased speed, and the conveyance carriages will advance. On a skip hoist the inductor will be rotated rapidly, and the hoist will accelerate under current limit in order to obtain maximum production. On a cage hoist, the inductor will be rotated more slowly and the hoist will accelerate under direct control of the inductor.

When the hoist reaches the correct point to start decelerating, solenoid M is energized to cause pawl N to strike level bar P, and stop the conveyance carriages. Further movement of the hoist will rotate the speed reference inductor back towards its creep-speed position. The hoist will always start to

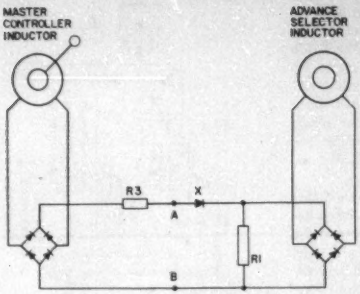
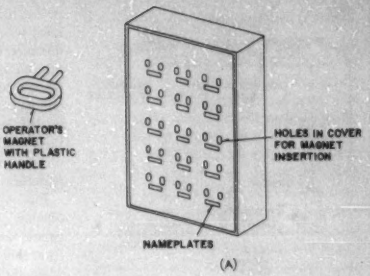


Fig. 9 (left). Minimum signal selector circuit



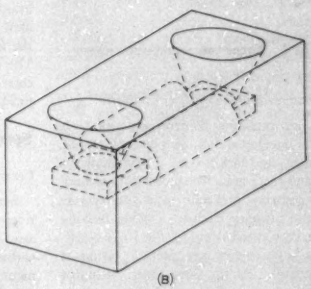


Fig. 10 (right). Contactless pushbutton

A—Station with operating magnet

B—Reactor encapsulated

decelerate from any speed at the right point to enable it to reach creep speed just before landing at the preselected level.

The advance selector is accessible, flexible, and easy to adjust. Additional hoisting levels can be readily accommodated by simply adding extra level bars to the equipment. It is suitable for application on a-c or d-c, friction, single- or double-drum hoists.

Overriding Speed Control for a Manual Hoist

One of the earliest steps towards automatic control of a Ward-Leonard hoist was the introduction of cam gear that operated through a "spring link" on the master control lever. The cams were profiled to return the control lever to the "off" position just before final stop was reached, and the operator had to exert pressure against the spring to bring the conveyance to the landing. This system was not entirely satisfactory in practice, because of the friction in the various levers and linkages.

Where the pattern signal is generated by a multipoint rheostat, contacts of the slow-down limit switches can be used to set up the circuits to the master controller contacts. In a typical scheme there would be 12 points on the master controller and four slow-down limit switches. Each limit switch would set up the circuit to the three speed points on the master controller.

The minimum signal selector circuit, developed for use with the advance selector and master controller inductors, eliminates the multiple contacts of the foregoing scheme. This circuit produces the pattern signal for automatic control, and provides overriding speed control when the hoist is manually operated. Fig. 9 shows the basic circuitry. The rectified output of the advance selector inductor is connected to resistance *R1*. The rectified output of the master controller inductor is connected via resistance *R3* and rectifier *X* to resistance *R1*, and the pattern signal is taken from points *A* and *B*. The ohmic value of *R3* is about 20 times that of *R1*.

Provided the master controller output is less than the advance selector signal, the operator will have full control of hoist speed, because the rectifier *X* will block any signal from the advance selector. If, however, the master controller signal exceeds that of the advance selector, the rectifier will conduct and the excess signal will be dropped across *R3*.

On automatic control, the master controller signal is replaced by a constant maximum-speed signal and the advance selector inductor then controls the hoist speed.

Underground Limit Switches and Pushbuttons[15]

The mine hoist rope is an elastic connection between the hoist drums and the conveyances. If accurate final stopping is to be achieved, the signal for stopping must be initiated by the shaft conveyance and not by the hoist drive.

Mechanical (lever-operated) limit switches in the shaft require constant maintenance. Magnetic proximity switches are preferable for this application. A permanent magnet, attached to the conveyance, operates a sealed relay located at the side of the shaft approximately 3 inches from the magnet. As the sealed relay involves contacts, a contactless, static proximity switch is being developed.

Conventional underground pushbuttons are another source of maintenance. On a 20-level semiautomatic cage hoist, for example, there will be over 450 underground pushbuttons. Any one pushbutton sticking could result in a shutdown.

A static control system has been developed in order to reduce this possibility of failure. Contactless pushbuttons are used in conjunction with Cypak static logic elements. The application of Cypak elements has a further advantage. They are phase-sensitive, and combinations of signals may be used to reduce the required number of shaft conductors. It is possible to transmit 15 signals, for example, on three signal conductors.

The contactless pushbutton is illustrated in Fig. 10. A small saturable reactor is encapsulated in a block of epoxy resin. The design is based on the principle of the inductance bridge, the balance of which is affected by an external magnetic field. The cage tender is provided with a specially shaped permanent

magnet. When this is placed in the locating holes of the encapsulating block, the reactor will produce an output signal. There are no moving parts, and only an operator with the correct magnet can initiate signals. Unauthorized tampering with such equipment is virtually impossible.

Automatic Hoist Safety Equipment[16,17]

Every hoist must be designed so that it will be brought to rest safely and an alarm given if abnormal conditions occur. The wide scope of this paper permits only a brief outline of some of the basic considerations in mine hoist operational safety.

In general, there are two kinds of abnormal conditions that may occur: those that must immediately initiate a hoist stop (overspeed, overtravel, etc.) and those that should prevent the hoist from starting again at the end of a trip (hot bearings, excessive brake wear, etc.). The former conditions generally are grouped together in the form of normally closed contacts in the safety circuit, while the latter are included in the alarm circuit. On manually operated hoists, alarm signals cause an alarm to sound; on automatic control, the alarm relay contacts are introduced into the safety circuit, but are by-passed by contacts of the directional contactors. For example, if a hot bearing develops during a hoisting cycle, the hoist completes the trip and then shuts down. Annunciators (such as glow tube triodes) usually are included in the safety and alarm circuits to facilitate the location of faults.

In English-speaking countries, the Lilly hoist controller is widely used as a means of overtravel and overspeed protection. This device, when properly adjusted, will give adequate protection as long as the slowdown cams are correctly profiled. It should be emphasized, at this point, that when an emergency trip occurs the brakes and only the brakes can bring the hoist to rest. The Lilly slowdown cams must be profiled with this consideration in mind, rather than "shadowing" the electrical deceleration. Various other systems have been evolved to check the hoist speed. Siemens of Germany uses a three position centrifugal switch to supervise hoist speed at three points of the slowdown period. ASEA of Sweden uses an a-c tachogenerator control system for the same purpose.

An emergency stop, by pushbutton, from within the cage is incorporated on some Canadian mine hoists by means of a fully transistorized transmitter and receiver which utilizes the hoist rope as a signal guide. Two-way voice communication, between cage and surface, is also possible.

Several new devices have been developed recently to facilitate the periodical safety checks on the performance of hoist brakes. Among these are

1. *Accelerometer.* An accelerometer is intended primarily to check deceleration during an emergency stop. It consists of a tachogenerator, driven by the hoist, and a differentiating circuit. The output of the differentiating circuit, which is proportional to the rate of change of speed, i.e., acceleration or deceleration, is indicated on a milliameter calibrated in feet per second[3].

2. *Distance-Measuring Device.* This device measures hoist travel as it comes to rest after an emergency stop signal is given. It consists of a pulsing device that gives, say, one pulse for every foot of rope travel, and an impulse counter that reads directly in feet.

3. *Brake Application Timer.* This device measures the time taken from the initiation of an emergency stop until the brakes are fully applied, and checks on the operation of the brake control valves.

Hoist Brakes[18]

It may seem out of place to discuss the mechanical hoist brakes in a paper on electric controls. On automatic hoists, however, the brakes must be considered as an integral part of the control. Most of the advances in the art of braking have been made by electrical manufacturing companies (the Metropolitan Vickers brake regulator and spring brake, the Siemens air-applied brake, the Canadian General Electric Microspot brake, etc.). The conventional mine hoist brake is of the caliper type and consists of a massive pair of brake shoes operating on a cylindrical brake race, as shown in Fig. 11. These brakes appeared very early in mine hoists, and have persisted almost unchanged until the present time. The brakes are applied by a dead weight, and released by air or oil pressure, and give adequate performance on a manually operated hoist.

On an automatic hoist, however, a fast-acting brake is essential if accurate stopping under all conditions of loading is to be achieved. The electric controls slow the hoist down to an almost constant creep speed. Power is then removed, and the brakes are applied. A fast-acting brake will reduce the distance traveled during braking, and so increase the landing accuracy under all conditions of loading.

If a conventional weight-applied brake is actuated rapidly, the inertia of the falling weight will cause the brakes to overshoot and "grab." The very high deceleration resulting from such an application will impose shock loads on the ropes and conveyances.

An obvious solution to the problem is to design a braking system that has a lower inertia. If the brakes are directly applied by air or oil pressure, the inertia of the dead weight is eliminated.

The Siemens brake engine (Fig. 12) is an excellent example of a modern air-applied brake. The brake shoes are directly applied by a large-diameter short-stroke piston, which acts at the middle point of a floating lever. In the event of an air failure, the backup weight

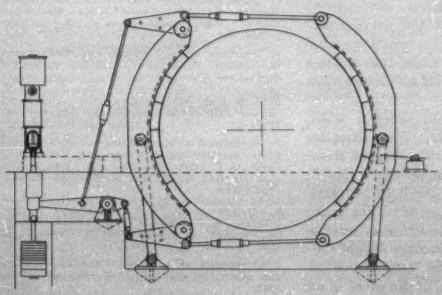


Fig. 11 (left). Conventional mine-hoist caliper brake

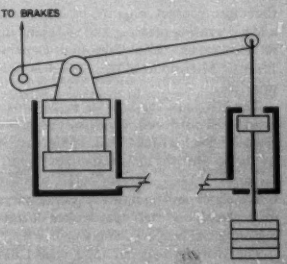


Fig. 12 (right). Siemens brake engine

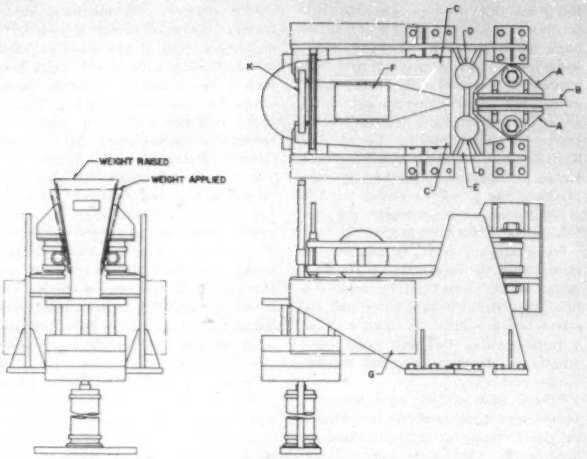


Fig. 13. Unit disk brake

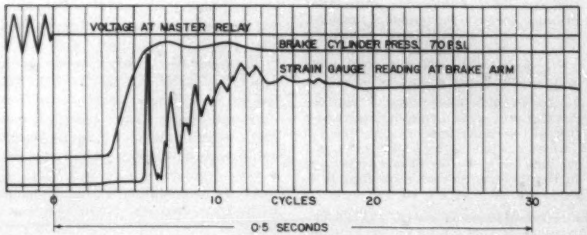


Fig. 14. Performance of unit disk brake

will fall and apply the brakes, as the point at which the air cylinder acts then becomes the fulcrum point.

The inertia of the massive conventional brake shoes still has to be overcome, however. Performance may be improved by first "priming" the brakes, i.e., bringing them into contact with the brake race at a low operating pressure and so taking up all clearances. The Canadian General Electric Microspot brake embodies this principle.

The inertia of the braking system may be radically reduced if a disk-type brake is used. Such a brake has been developed by both British and Canadian manufacturers. The conventional massive brake shoes and cylindrical brake races are replaced by small spot brakes operating on radial braking flanges or disks.

The British National Coal Board design is based on aircraft practice; the brakes are spring-applied and released by very-high-pressure oil (1,500 pounds per square inch). The Canadian Westinghouse design is air-applied, with an emergency weight backup as shown in Fig. 13. This brake is intended primarily for Koepe hoists. On a typical installation, there will be two brake disks and two brakes per disk, i.e., four independent brakes (see Fig. 13).

Each brake has two shoes A which clamp onto the disk B. These shoes are attached to levers C which pivot about pins D located in a supporting frame E, which is mounted on the hoist bedplate. The brake is applied by admitting air to the center of a double-acting cylinder F to force the levers C apart. A light tension spring releases the brakes when air pressure is removed. In the event of an air failure, or maloperation of the control valves, a backup weight W will fall. A wedge K attached to this weight will force the brake levers apart.

The whole system is self-aligning and able to follow irregularities or waviness in the brake flange. All braking components are mounted on or above the hoist bedplate, thus simplifying installation and routine maintenance.

Each brake is designed as a unit, and is fitted with its own set of control valves. Braking effort is controlled by one set of master pressure-regulating valves. It is determined by operating pressures, not by a position follow-up system as on the older conventional brakes, so that no adjustments of linkages are necessary.

As the inertia of the disk brake is much lower than that of the caliper brake, very rapid applications without overshoot are possible. The performance curve of the brake is shown in Fig. 14. Full brake pressure can be built up in the operating cylinder in 0.1 second and the brake is fully applied, with negligible overshoot, in 0.15 second, without the necessity of priming.

Conclusions

This paper has described trends in automatic mine hoist controls, and shown that considerable progress has been made recently in this field. The future will bring more widespread use of fully automatic skip hoists, when problems involved in automatic loading have been solved.

On cage hoists, the accuracy of high-speed passenger elevator control will soon be matched.

References

1. Electric Winders, H. H. Broughton. Spon, second edition, 1948.

2. Winding Engine Calculations for the Mining Engineer, A. B. Price. Fraser & Chalmers Engineering Works, Erith, England, May 1955.

3. Some Aspects of Electric Winders, H. M. Hughes. *The Mining Electrical (and Mechanical) Engineer*, Manchester, England, vol. 36, May 1956, p. 299.

4. Friction Hoist Design Considerations, G. L. Tiley. *Transactions*, Canadian Institute of Mining and Metallurgy and of the Mining Society of Nova Scotia, Montreal, Que., Canada, vol. LX, 1957, p. 176; also *Canadian Mining and Metallurgical Bulletin*, Montreal, Que., Canada, June 1957.

5. Symposium on Electrification of Dominion Coal Company Steam Hoists, J. A. Russel, J. J. Laffin, G. L. Tiley. *Ibid.*, vol. LXI, 1958, p. 194; also *Canadian Mining and Metallurgical Bulletin*, June 1958.

6. A Review of Modern Electric Winder Control, E. Friedlander. *The Mining Electrical (and Mechanical) Engineer*, vol. 28, Dec. 1957, p. 159.

7. Symposium on Automatic Hoisting. Glückauf, Essen, W. Germany, Dec. 1957, vol. 93, p. 1517.

8. D-C Winder Control Using Mercury Arc Rectifier, L. Abram. *The Metropolitan-Vickers Gazette*, Manchester, England, vol. 28, July 1957, p. 156.

9. Application of Dynamic Braking to A-C Mine Winders, O. T. Evans, L. Abram. *Ibid.*, vol. 25, no. 410, Sept. 1953, p. 48.

10. Low-Frequency Performance of a Wound-Rotor Induction Motor for Mine Hoist Drives, F. F. Dixon, G. L. Tiley. *AIEE Transactions*, vol. 76, pt. III, Dec. 1957, pp. 1140–45.

11. A Low-Frequency Generator and Frequency Changer for A-C Mine Hoist Drives, C. W. Elliot, W. Boesel, E. Zucker. *Ibid.*, pp. 1073–77.

12. Reactor Control of A-C Winders, H. Takeda, D. Moriya. *Report. 340 HF Hoist.* Mitsubishi Mining Company, Oyuban Mine, Japan.

13. Load-O-Matic A-C Crane Hoist Control, W. R. Wickerham. *Westinghouse Engineer*, Pittsburgh, Pa., Mar. 1952, p. 56.

14. Friction Drive Mine Hoists—Recent Installations and Design Considerations, E. P. Pfleider, E. G. Malmlow, F. Landow. *Mining Engineering*, New York, N. Y., Jan. 1955.

15. Magnetic Logic Circuits for Industrial Control Systems, W. G. Evans, W. G. Hall, R. I. Van Nice. *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 166–71.

16. Putting the Type "C" Lilly into Service, G. L. Tiley. Application Engineering Publication Service, Canadian Westinghouse Company, Ltd., Hamilton, Ont., Canada, Dec. 1957.

17. Safety of Equipment for Electric Winders, W. Walker. *The Metropolitan-Vickers Gazette*, Aug. 1956, p. 252.

18. Early Developments of a New Mechanical Braking System for Winders, N. Dudley. *The Mining Electrical (and Mechanical) Engineer*, vol. 37, June 1957, p. 331.

19. Trends in Design of Tower Mounted and Ground Mounted Friction Winders, P. H. Harvey, E. Klebert. *Ibid.*, vol. 28, Mar. 1958, p. 247.

Discussion

J. G. Ivy (International Minerals and Chemical Corporation, Carlsbad, N. Mex.): This paper represents an important contribution to the literature on electric drives for mine hoists.

First, it represents an important summary of the "state of the art" in 1958. Viewed in retrospect by future generations this summary will become important history.

Second, and perhaps of even greater importance, this paper gives a concise summary of the alternatives available to the designer of a hoisting installation. The hoist, drive, conveyance, headframe, and shaft combination, considered together as a machine, is the most important and also the most expensive single piece of "equipment" in any mine. Good design, in the selection of the components of this machine and the establishment of their proper proportions, can pay large returns in the realization of adequate performance for minimum capital investment. This design is usually not solely the work of the electrical engineers, but rather is a team effort of electrical, mechanical, and mining engineers. The paper has great worth in assisting the electrical member of this team in presenting the electrical aspects of the over-all hoist design to his nonelectrical teammates.

P. deH. Eastcott (Canadian General Electric Company, Peterborough, Ont., Canada): The authors of this paper are to be complimented on their objective appraisal of the technique and equipment being used in modern North American hoisting installations. Since space has dictated that their treatment of the subject be of a general nature, the following amplifying remarks may be of interest.

Regulators

Although not mentioned by the authors, acceleration control in addition to accurate speed regulation is a highly desirable feature, particularly on automatic friction hoists. During the last 4 years, all d-c adjustable-voltage mine hoist drives supplied by Canadian General Electric have employed acceleration-rate limit control. At the same time speed regulation as a function of generator voltage with *IR* compensation was abandoned in favor of the more precise direct regulation of hoisting speed by means of tachometer generator voltage feedback.

Although several types of acceleration-rate feedback control are possible, it is generally accepted that the simplest and most reliable system employs an auxiliary generator field winding which, through transformer action with the main shunt field, provides a voltage signal proportional to rate of flux change. It is a most fortunate property of armature-voltage-controlled d-c drives that the first derivative with respect to time of generator flux, generator voltage, motor speed, and consequently acceleration are all approximately equal. Therefore, if the appropriate amount of rate field signal reaches the regulator as negative feedback, an accurate limit of hoist acceleration results, independent of conveyance load or of whether speed is increasing or decreasing.

Acceleration-controlled hoist drives are markedly smoother in operation than their counterparts which rely solely upon current-limit feedback.

As an incidental function, generator-rate feedback imparts a large measure of stability to the high-gain speed regulator of which it forms a part, and simultaneously improves motor and generator commutation through reducing the magnitude, rate of change, and the number of armature-loop current swings during normal hoisting and spotting operations.

Program Devices

Sound hoist-design policy dictates that manual control, used in such emergencies as underground fires, employ the bare minimum of control equipment. Consequently, it is advisable to avoid the use of pilot motors, limit switches, slip clutches, and differential gears in program switch design, thereby keeping to as simple and rugged construction as possible. Where acceleration and current (torque) limit controls are incorporated in the regulator and are thus available during both manual and automatic operation, there is little to be gained by deluxe program equipment providing more than two or three accurate switching points per slowdown pattern. Conversely, where a pilot motor provides the acceleration function by changing the speed pattern signal or reference at a constant rate, as in the advance selector described, the acceleration limit feature is usually inoperative during manual control. For use on a-c automatic drives, and on d-c drives where the ultimate in control accuracy is required, Canadian General Electric has in service a multilevel program switch which provides continuous slowdown profiles for up to 12 levels. Each independent cam unit is the equivalent of a single-turn cam 40 feet in circumference, thus insuring accurate approach signals for each level. Where necessary, a number of 12-stage switches may be operated in tandem.

Brake Test Equipment

The electric accelerometer, distance-measuring device, and brake application timer, of which many types are available, have been widely adopted in Canada during the past 3 years. Considerable experience has been accumulated in the use of these devices from which the distance-measuring device appears to be the most useful of the three for daily or routine brake testing once the safe limits have been established. Mechanical pulse counters are not used by Canadian General Electric for stopping distance measurements because most commercial counters are subject to stuttering (missing counts at high speed) if the transmitter contacts are not kept in good adjustment. Should stuttering occur, the actual stopping distance will be greater than that indicated, and a corresponding reduction in brake effort may lead to a serious accident. The brake application timer has not proved to be of much value on installations based on the Microspot brake design. With three or more dump valves operating in parallel and properly proportioned piping, the failure of any two valves seldom increases brake application time by more than 25%. The chance of three simultaneous valve failures occurring with statutory daily inspections is hardly comparable with the normal hazards of underground mining.

Brake Design

The authors imply that the ability to "prime" a braking system (remove running brake shoe clearance) is a crutch to overcome poor design but, in fact, priming will bring out the best of any braking system for the cost of a single magnet valve. Priming reduces the normal stop application time to somewhere between 1/2 to 1/20 of the emergency setting time, which facilitates the accurate spotting of a conveyance. Of more significance is the ability of priming to minimize impact stresses in the brake mechanism and suspended system.

Primed and emergency stop brake applications must be considered separately when designing hoist brakes. It is possible to build a brake which is quite satisfactory from the standpoint of primed applications, but which fails miserably due to inertia overshoot under emergency stop conditions.

The authors state that through enlightened design which has eliminated the massive shoes and rods of the parallel motion brake, the Westinghouse disk brake does not require priming and may be applied without overshoot in 0.15 second. The massive shoes and main pull rods of a modern pressure-applied braking system as

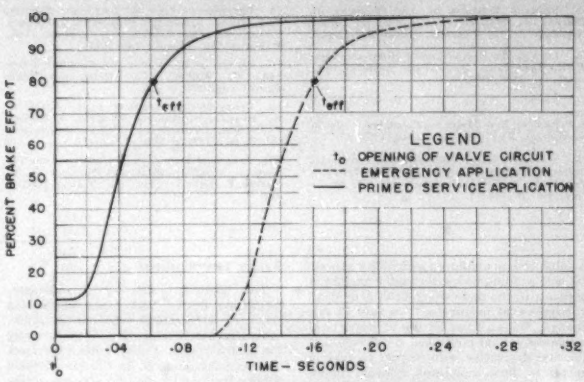


Fig. 15. Microspot air brake characteristics

Time–brake effort curve for 10-foot parallel motion hoist brake

built in North America do not store more than 4% of the impact energy during a rapid unregulated emergency application. Cutting this down to 2% is hardly significant. Large disks or extended flanges with their weight located at a greater radius than the conventional brake paths which they replace, do not necessarily result in reduced hoist inertia, particularly if the equivalent capacity for heat storage is provided. However, the disk with its large unswept areas and radial air scrubbing is a more efficient energy converter. This feature alone may justify its adoption but it will be interesting to observe the day-to-day stability of the vertical flange or disk brake on friction hoists located over wet shafts. Fines, water, oil, and rope dressing already present problems by finding their way onto conventional brake paths enjoying the slinger protection of substantial drum flanges. Although slinger rings and covers may be used with the disk brake in an effort to keep the disks dry and clean, many designs are impractical during sinking operations where dressed ropes and shaft water may both be encountered.

On the other hand, the parallel motion brake does not have ease of adjustment in its favor.

Brake Performance

In comparing the results obtained during a large number of field and shop tests on both pressure-applied and falling-weight hoist brakes, it has been found useful to introduce the concept of effective setting time, t_{eff}, as being numerically equal to the delay in seconds measured from the initiating signal to the time at which the brake effort reaches 80% of its final steady-state value. The figure of 80% has been arbitrarily selected since it represents a substantial fraction of final steady-state braking while at the same time eliminating the top end of the effort versus time characteristic which is not of great significance unless overshoot is present. It is hardly fair to penalize a brake with a well rounded time–effort characteristic, in itself highly desirable, merely because it took the better part of half a second for the final 5% of brake effort to be applied.

By way of comparison, the following curves of brake effort against time have been obtained during testing of parallel motion air brakes employing brake engines based upon the Microspot design; see Fig. 15.

Brake effort was measured by means of 2,000-cycle electric strain gages mounted on the main pull rods, and recorded by photographing an oscilloscope trace of the resulting rod strain. At time t_o, current to the brake valves was interrupted. Where the brake had been previously primed, an effective setting time of 0.061 second was obtained; while for a fully released emergency application the effective setting time increases to 0.16 second. In both cases, a well-rounded characteristic free from overshoot results.

An Automatic Speed-Control System for a Gravity Freight-Classification Yard

R. J. BERTI
ASSOCIATE MEMBER AIEE

T. J. DOSCH
NONMEMBER AIEE

IN RECENT YEARS claims for damage to freight shipments have been steadily increasing, and such claims currently cost the railroads of the United States over 100 million dollars annually. Some of this damage occurs during the various classification processes. There are two main types of classification yards; namely 1. flat switching and 2. gravity. The gravity yard is the type of yard dealt with in this paper.

Manual operation of gravity-retarder yards is complicated by the dual requirements of maintaining a smooth, continuous flow of traffic and of avoiding the high coupling speeds which cause impact damage. Under pressure of heavy movement of traffic, the operator may not have sufficient time to give proper attention to each car and this may result in some couplings occurring at a higher than the desired speed.

The advent of electronic measuring and computing equipment has made practical the development of automatic retarder yards. Although some labor reduction may result, such automatic operation must reduce damage claims to be economically justifiable. This paper discusses some of the factors which affect coupling speeds, and describes an automated computer system installed by Reeves Instrument Corporation at North Platte, Neb.

In the summer of 1953 preliminary data on the rolling characteristics of free-rolling freight cars were obtained at the North Platte retarder yard, and a contract for the development of a prototype control system was entered into between the Union Pacific Railroad and Reeves Instrument Corporation of New York City. This prototype system included standard d-c analog computer components, and provided for control of one group of eight tracks. The result of this

Paper 58-1266, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 24, 1958; made available for printing September 4, 1958.

R. J. Berti is with the Union Pacific Railroad Company, Omaha, Neb.; and T. J. Dosch is with the Reeves Instrument Corporation, Garden City, N. Y.

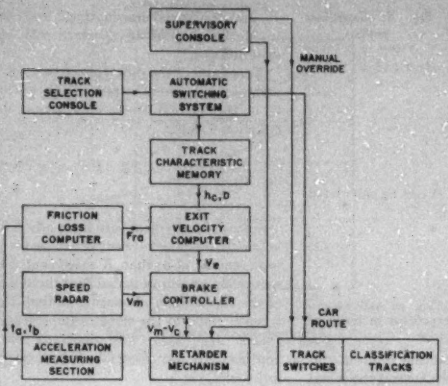


Fig. 1 (left). System block diagram

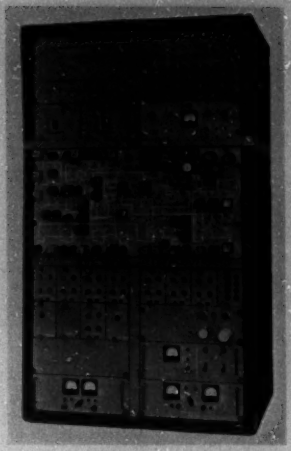


Fig. 3 (right). Retarder controller cabinet

prototype installation proved that full-scale automation was feasible and subsequently the entire yard has been placed under automatic control. Since then a second computing system has been installed at the Southern Railway Company's Inman yard in Atlanta, Ga.

Nature of the Problem

The basic layout and nature of a gravity-type classification yard have been previously described and the control system discussed is for use in such a yard. Fig. 1 is a block diagram, and Figs. 2, 3, and 4 are views of the equipment discussed in this paper.

Briefly, the control system utilizes a computer which determines the acceleration of each car on a tangent track of known slope. The cars pass through this "acceleration measuring section" of track before entering the group retarder. From knowledge of each car's acceleration on this known slope, and from other factors to be described later, the computer determines the speed at which each car should leave the group retarder in order to couple properly. A closed loop then controls each car to this computed speed as it passes through the group retarder. The system also includes a switching machine for the automatic routing of cars. As will be seen, the switching machine is a necessary part of the speed-control system because the outputs from the switching machine are used to select the track data appropriate to a given car destination.

Before describing in more detail the computer used for the speed-control system, some basic factors that affect the motion of cars will be reviewed.

System Equations

A simplified profile of the route of a car from its point of release from the group retarder to its point of coupling is shown in Fig. 5. Considering the motion of the car between points a and c, it can be shown that

$$V_a^2 + K_1\left(\frac{v_a}{r}\right)^2 = v_t^2 + K_1\left(\frac{v_t}{r}\right)^2 + 2gD(F_{ra} - M_a) - 2gh_t + 2gd_tF_c \quad (1)$$

where

v_a, v_t = speed of the car at points a and c respectively
r = radius of the car wheels
K_1 = ratio of the moments of inertia of the wheels and axles of the car to the car mass
F_{ra} = average frictional force acting on the car on tangent track expressed in terms of the equivalent slope, i.e., a slope on which the car would roll at constant velocity
F_c = average frictional force acting on the car due to track curvature and turnouts, expressed in terms of the equivalent slope
= gravitational acceleration

Fig. 2. General view of equipment room, North Platte

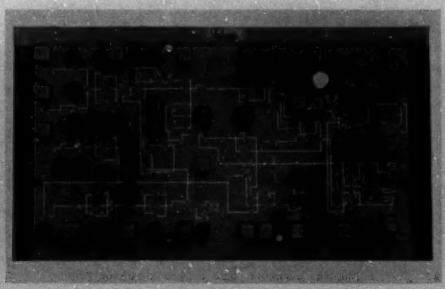


Fig. 4. Retarder controller functional front panel

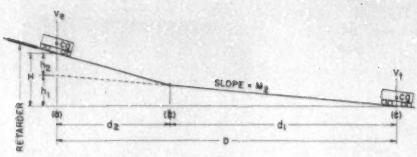


Fig. 5. Simplified profile

The other quantities are defined in Fig. 5.

Let

$$1+\frac{K_t}{r^2}=\frac{1}{K}$$

and let $d_2F_c-h_2\times h_c$

Equation 1 then becomes

$$\frac{v_e^2}{2g}=\frac{V_t^2}{2g}+KD(F_{ra}-M_2)+h_c \qquad (2)$$

Note that the terms of equation 2 have the dimensions of distance. As in hydraulic systems, it is useful to consider the various energy components in terms of an equivalent potential energy head, and this is done throughout this paper. For example, a car which is rolling at 5.5 mph (miles per hour) is said to have approximately one foot of velocity head.

The quantity K is seen to be a measure of the ratio of the rotational energy stored in the wheels and axles of a car to the total energy of the car. Since for a 20-ton empty car the rotational energy is about 10% of the total energy, K can range in value from about 0.90 for an empty car, to 0.98 for a car and load with a total weight of 100 tons.

The factor K must also be considered when calculating the frictional force acting on a car from a measurement of its acceleration on a known slope M_1; see Fig. 6. From the expression for the motion of a car on this known slope, an equation from which F_r can be calculated is obtained.

$$F_r=M_1-\frac{a}{gK} \qquad (3)$$

where

F_r = frictional force acting on the car on tangent track expressed in terms of the equivalent slope

It will be seen that the error in calculating F_r can be very large if the wheel energy is neglected, i.e., if K is taken to be equal to unity. For example, a 20-ton car with a true F_r of 0.3% would yield a calculated F_r of nearly 0.5% if the measurement was made on a 2% downward slope, and the wheel energy was neglected.

An expression for the error in the calculation of the required group-retarder exit velocity due to an error in K will now be derived. (It is assumed in this derivation that friction is independent of velocity, so that $F_{ra}=F_r$.) The value of F_r so obtained will be defined as F_r', as given by the expression:

$$F_r'=M_1-\frac{a}{gK'} \qquad (4)$$

where K' is the assumed value of K.

From equations 3 and 4, F_r' can be written as a function of F_r as follows:

$$F_r'=M_1-\frac{K}{K'}(M_1-F_r) \qquad (5)$$

Substitution of this value of F_r' for F_r, and of K' for K in equation 2 leads to a calculated value of v_e which will be designed as v_e'.

A measure of the difference between v_e' and the proper group-retarder exit velocity v_e is then found from substitution in equation 2 to be:

$$\frac{v_e^2}{2g}-\frac{v_e'^2}{2g}=D(M_1-M_2)(K-K')+h_c(K-K') \qquad (6)$$

Since h_c in a typical yard is usually not larger than about 2.0 feet, it can be seen from equation 6 that K need not be accurately known provided that the slope of the acceleration measuring section M_1 is made equal to the slope of the storage tracks M_2. This was done at the Southern Railway System's Inman yard in Atlanta, Ga.

Curved-Track Resistance

The computer to be described assumes that all cars traveling to a particular track experience the same h_c head loss. Actually, the curved-track loss varies from car to car, and even varies considerably for the same car on successive runs over a given section of curved track. Fig. 7 shows the velocity-head loss experienced by a standard test car (to be described in following paragraphs) during 33 consecutive runs on track no. 37 at North Platte, and the head loss for 48 random cars on track no. 30. It will be realized that the deviations of curved-track loss from the mode are large enough to cause considerable dispersion in coupling speeds.

Although measurements of curved-track resistance for each car can be made, the value of such measurements is limited in many yards. As examples, skewed trucks may cause a low curve loss for one direction of turn, and a high loss for the opposite direction, and brakes which

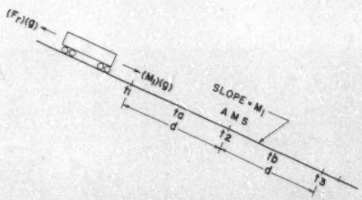


Fig. 6. Acceleration-measuring section

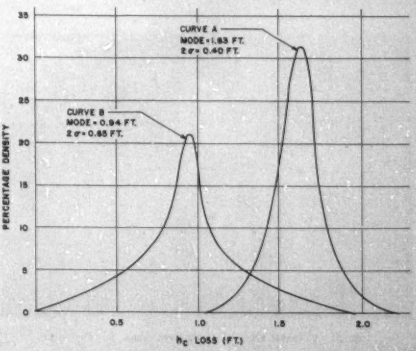


Fig. 7 (right). Typical h_c losses, North Platte

Curve A—33 runs with test car, track no. 37, 2 turnouts + 13 degrees 33 minutes curvature, 550 feet to clearance point

Curve B—48 runs with random cars, track no. 30, 2 turnouts + no curvature, 320 feet to clearance point

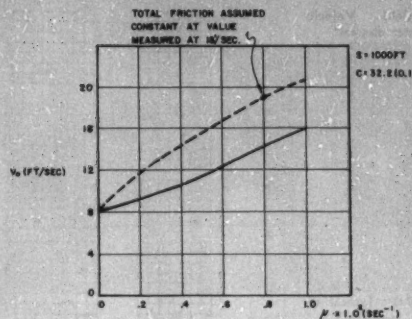


Fig. 8. Effect of viscous drag on V_0

Fig. 9. Test-car pulse generator

appear free when a car is on tangent track or on a mild curve may tighten when the same car negotiates a sharper curve due to the eccentric placing of the brake-rod linkages.

Therefore, if the degree and direction of curvature are not the same before and after the group retarders, and this is usually the case, it is questionable whether measurements on each car of curve friction made prior to the group retarder are useful in determining a car's motion on the curves after the group retarder. It is for these reasons that the added complexity of curved-track resistance measurements for each car were not felt to be justified in the systems installed at the North Platte and Inman yards.

Viscous Friction

As is well known, the F_r of a car increases with increasing velocity. To a first approximation the frictional force on a car may be represented as the sum of a constant force (coulomb friction) and a force proportional to velocity (viscous drag). The equation of motion for such a car on a slope M_s is

$$M_s mg - f_s - u_c v = m\frac{dv}{dt} \qquad (7)$$

where

f_s = the coulomb friction
u_c = the coefficient of viscous friction, and
v = the velocity of the car

From equation 7 an expression can be obtained which relates the initial velocity to the distance traveled:

$$CS + \frac{C}{u}(v_t - v_0) + \frac{C^2}{u^2}\log\frac{uv_0 + C}{uv_t + C} = 0 \qquad (8)$$

where S is the distance which the car travels in slowing from v_0 to v_t

$$C = \frac{f_s}{m} - M_s g$$

$$u = \frac{u_c}{m}$$

v_0 = initial velocity

For $S = 1{,}000$ feet, Fig. 8 shows the initial velocity as computed from equation 8 for various cars. Fig. 8 also shows the exit velocity which would be computed if F_r were assumed constant, and equal in value to the friction measured at 18 feet per second (fps) (the value of 18 fps was chosen since this is approximately the speed at which cars pass through the acceleration-measuring section in the North Platte yard). As an example, consider that the F_r of an empty car is measured at 0.5% while traveling at 18 fps, and that F_r approaches 0.1% at zero speed. This ignores the increase in friction in the region near zero speed. In order to roll 1,000 feet ($v_t = 0$) the initial velocity must be 13.4 fps. If the viscous-drag effect was ignored, and the measured F_r value of 0.5% was used in equation 2, the car would have been given an initial velocity of 17.9 fps. A correction circuit for viscous drag which included the effects of distance to roll was first designed

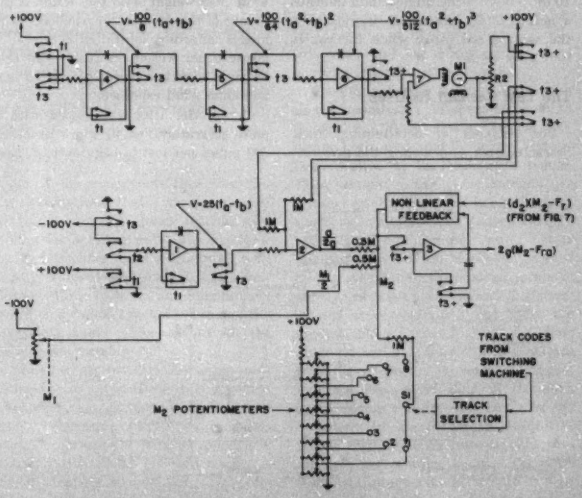


Fig. 10. Friction-loss computer diagram

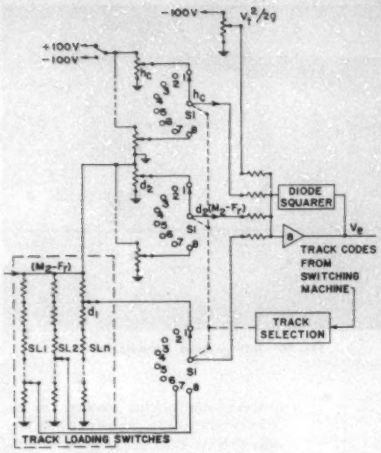


Fig. 11 (left). Velocity computer diagram

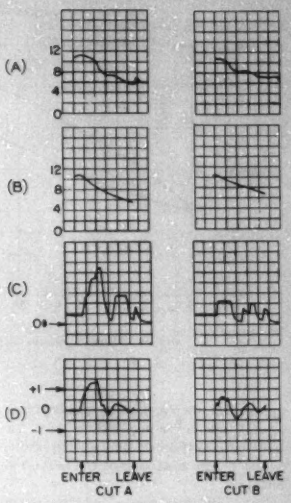


Fig. 13 (right). Typical retarder operation

A—Radar speed, V_m, miles per hour
B—Called-for speed, V_e, miles per hour
C—Air pressure, p
D—Error, $V_m - V_e$, miles per hour

for and installed at the Southern Railway's Inman yard, and was later included in the North Platte installation to be described in a following section.

It has been found that the increase in F_r with velocity is far larger for cars rolling singly, than for the same cars when coupled in a train. For example, typical data for cars in a train show an increase of about 10% in the friction of a 25-ton car when the car speed is increased from 10 to 15 mph, while limited data indicate a nearly 100% increase in friction over the same speed range when the car is rolling by itself.

The Test Car and Its Uses

For purposes of determining track characteristics, and testing the performance of the system in general, a test car was constructed which was as nearly an ideal car as was felt to be practicably attainable.

The test car consists of a 40-foot flat car incorporating the following: all roller-bearing journals, perfectly aligned sized and coned wheels, bronze center bearings, proper side-bearing clearances, and springs attached to brake shoes to prevent dragging. The car is weighted with rail to a total gross weight of 55 tons which is the weight of the average car handled. The load is uniformly distributed on the car to insure equal weight on center and side bearings with low center of gravity and minimum wind exposure.

One of the axles is equipped with a pulse generator (Fig. 9) which produces 200 pulses per foot of car travel. These pulses modulate a transmitter on the car for telemetering the car's speed to a central equipment room. The transmitter has a range of approximately one mile by means of which it is possible to accurately monitor the car's behavior at all points in the yard. After being received in the equipment room and demodulated, the pulses are fed to both a preset counter and an events-per-unit-time (EPUT) counter. After each count of 4,000 pulses, i.e., 20 feet of car travel, the preset counter initiates a digital print-out of the car's speed as indicated by the EPUT meter. The rate of velocity-head loss suffered by the car at all times is thus easily determined, and deducting from this the known friction loss characteristic of the test car, the track characteristics are readily and accurately obtained. In addition, test runs can easily be made periodically to insure that the speed control system stays in calibration.

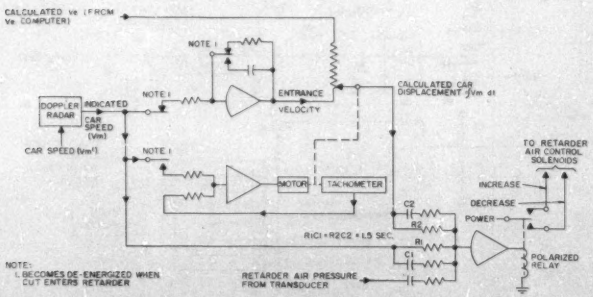


Fig. 12. Retarder controller diagram

Control System Components

Friction Loss Computer

The most exacting computation required of the system is the determination of F_r. This is readily apparent when it is considered that to achieve couplings to an accuracy of 0.6 foot of head at a range of 2,000 feet requires a measurement of acceleration to an accuracy of (0.0003) (g). The times of contact of a leading car wheel with three equally spaced insulated track circuits located on tangent

Fig. 14. Radar antennas, North Platte

track of uniform grade are used for this purpose as shown in Fig. 6. As shown in the figure, t_a and t_b represent respectively the times required by the car wheel to travel between the first two and the last two insulated track joints. The computation of acceleration is considerably simplified because t_a is approximately equal to t_b. It can be shown for this case that negligible error is incurred by using the approximation:

$$a \approx 8d \frac{t_a - t_b}{(t_a + t_b)^3} \tag{9}$$

The solution of equation 9 is accomplished by the computer organization shown in Fig. 10. The time notation at each relay in Fig. 10 refers to the time at which the relay is energized by track circuits t_1, t_2, and t_3 (see Fig. 6). The terms of the equation are computed as follows: $(t_a - t_b)$ is obtained by applying a positive voltage to the input of integrating amplifier no. 1 during the interval t_a, and an equal but opposite voltage to the amplifier during interval t_b. The $(t_a + t_b)^3$ term is generated by means of a triple integration by means of amplifiers 4, 5, and 6. A servomotor driven by amplifier no. 7 positions potentiometer $R2$ so that its output is equal to $(t_a + t_b)^3$. At time t_3 this potentiometer is switched into the feedback circuit of amplifier no. 2, thereby producing the division of $(t_a - t_b)$ by $(t_a + t_b)^3$. The output of amplifier no. 2 is thus proportional to the acceleration.

The frictional loss F_r is then obtained by subtracting $a/2g$ from $M_{1/2}$ at the input of amplifier no. 3. At this point an average wheel energy factor is set into the equation by altering input $M_{1/2}$.

The F_r which has been computed at this point is, of course, not the "average" F_r (i.e., F_{ra}) specified in equation 1 but is the F_r which the car exhibits at the speed at which it passes through the acceleration-measuring section. Since the friction of the car has been measured at only one speed, a solution cannot be obtained for the average F_r without making an assumption concerning the variation of friction with velocity. Further, the distance to roll modifies the average F_r. The necessary corrections are approximated by the nonlinear feedback indicated in Fig. 10.

It is, of course, true that a car with dragging brakes or any similar fault will not have a proper F_{ra} computed because the ratio of coulomb friction to viscous drag for such a car is abnormal.

By subtracting the classification track slope M_2 from F_r at the input of amplifier 3, the output can be made proportional to the velocity head loss of a given car over a given length of track. The M_2 slopes proper to the track for which the computation is being made is selected by stepping switch S_1 (Fig. 10) which in turn is positioned by the automatic-switching system.

The output $2g\,(M_2 - F_r)$ is subsequently stored in amplifier no. 3, and is applied as one of the input factors to the velocity computer.

Velocity Computer

All of the components that comprise the velocity head that a car should have at the exit of the group retarder are summed up as shown in Fig. 11. The feedback of amplifier no. 8 is a network consisting of biased diodes which act to produce a voltage from the amplifier proportional to the square root of the sum of the amplifier inputs. The output of the amplifier is therefore v_s.

The stepping switches shown are positioned by the automatic-switching system for each car to correspond with the proper track destination. Input h_s is determined from test-car data as previously noted, and is preset for each track. A single stepping switch *SL1*, *SL2*, and *SLN* is associated with each track. The position of the arm of each stepping switch represents the number of cars that have already been stored on the associated track. These are initially set by a supervisory operator in a control tower and subsequently automatically stepped one increment for each car.

Retarder Controller

Having computed the velocity at which the car should leave the final retarder, the car must now be acted upon by the final retarder in order that the car leave at the calculated v_s. This is accomplished as shown in Fig. 12. A programmed speed is computed which is made equal to the car's speed at the entrance to the retarder, and is gradually reduced until it is equal to the desired exit velocity when the car leaves the retarder. The speed of the car is continuously monitored by means of doppler-type radar, and the speed so indicated is compared continuously with the programmed speed. The difference is used to actuate a relay which in turn increases or decreases the retarder pressure as required.

The Union Pacific North Platte yard uses pneumatically operated retarders with a time constant of approximately 1.5 seconds. Position feedback from a pressure-voltage transducer connected into the retarder air chambers acts to linearize the on-off response. Fig. 13 shows typical recordings of air pressure P, programmed speed V_c, radar speed V_m, and error betweeen programmed and actual car speed $V_m - V_c$. The brake error has an rms value of approximately 0.3 mile per hour, and the effect on the final coupling error is therefore small.

The original radar-antenna installation used in monitoring car speed consisted of 8-element dipole arrays having 20-degree horizontal, and 40-degree vertical beam widths at the half-power points. These antennas, using one antenna per retarder group, were tried in a number of locations, none of which proved entirely satisfactory. Optimum results were finally obtained in the North Platte yard by changing to a 4-foot-diameter parabolic antenna having a beam width of 7 degrees, located approximately 3 feet above the ground plane; see Fig. 14. These antennas are located 60 feet upstream from the entrance to the retarder and provide excellent monitoring of the car's speed during the time the car is in the retarder. In the Southern Railroad's Inman yard, three separate dipole arrays are used per retarder group. These are summed through a diode OR type of logic circuit. Since the output of the OR cir-

cuit is always the largest of the three radar speeds, proper placement of the antennas insures a continuous-speed voltage so long as at least one of the radars has a clear signal.

Range Determination

Some means is required to enable the operator to know the loading on the class tracks, some of which at times are out of sight. To this end a ranging system was devised by means of which a supervisory operator can interrogate any track at will to determine the amount of space remaining between the coupling point and the clearance point.

When a track is interrogated a source of 60-cycle constant current is connected at insulated joints located at the clearance point of the particular track being interrogated. The current amplitude is such as to produce one millivolt of drop for each foot of range to the coupling point. Thus, when the nearest car is 900 feet from the clearance point a voltage of 900 millivolts is produced at the beginning of the track. This voltage is brought back to the equipment room where it is amplified, converted to direct current, and fed to an analog to digital converter designed to produce an indication of track loading at the supervisory console. The range setting in the computer can be brought to correspondence with the situation on the track, or changed as otherwise may be desired, by means of appropriate push-buttons.

The operator has the option of interrogating tracks individually as described, or by depressing a button marked "Automatic" he can set into motion an automatic scanning device which sequentially scans each track in the yard. Since scanning a track while cars are in motion can produce erroneous range settings, the operator need only watch that the automatic scanning is not in service during the humping operation.

Conclusions

There are several factors pertaining to yard design that tend to improve results:

1. Construct the acceleration-measuring section as nearly as practicable to the same slope as the classification tracks to reduce the rotational energy-error factor.

2. Design classification groups with as few tracks as consistent with economy, to keep the switching leads as short and free of curves and turnouts as possible. The more nearly that tangency is maintained beyond the group retarders, the less the dispersion in h_s from the mode value set into the system.

3. Design the gradients of the switching leads to cause the average velocity of the car over these sections to be as nearly as possible the same as while in the acceleration-measuring section. This minimizes the effect of viscous drag.

The basic requirements for the successful automation of a gravity-type freight-classification yard to reduce impact damages are now adequately known. The computations required lend themselves readily to solution by d-c type analog computing elements. Numerous forms of system organization to meet these requirements are possible, and a great variety of system components are available.[2]

Several yards have been automated that demonstrate the validity of the concepts discussed. It is to be expected that further improvements in automatic control will continue to be made, and that ultimately manually controlled freight-classification yards will be entirely replaced by automatic yards.

References

1. Automation for Gravity Freight Classification Yards, A. V. Dasburg. *AIEE Transactions*, vol. 74, pt. II, Nov. 1955, pp. 320-24.

2. Developing the Automatic Retarder Controller, Perry A. Seay. *Proceedings*, Railway Systems and Procedures Association, Arlington, Va., Spring 1955.

The Application of Shift-Register Techniques to Materials Handling

HENRY C. DIENER, JR.
ASSOCIATE MEMBER AIEE

Synopsis: Shift registers provide a versatile building block for control of industrial materials handling systems. When composed of static elements, flexibility and reliability are combined in performing control functions. Circuit configurations and field applications are described for dispatching, scanning, counting, and static stepping switch functions.

THE TERM "shift register" is well known to those familiar with digital computer circuitry.[1] However, today it is finding wide application in static-type industrial control systems as well. Evans, Hall, and Van Nice have described industrial-type magnetic circuits for performing the basic English logic functions of AND, OR, NOT, MEMORY, and DELAY.[2] The shift-register circuit applications described in this paper are composed of this magnetic logic system. Such systems and logic elements, employing the Ramey magnetic-amplifier principle, are commercially available under the trade name of Cypak.[3] However, the principles to be described are sufficiently general to apply to other forms of static control as well.

Paper 58-1166, recommended by the AIEE Industrial Control Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted June 27, 1958; made available for printing July 30, 1958.

Henry C. Diener, Jr., is with Westinghouse Electric Corporation, Buffalo, N. Y.

Fig. 1 illustrates the basic shift-register function. Here are shown five shift-register stages in cascade. Using conventional binary notation, *1* designating ON and *0* designating OFF, the five stages are arbitrarily assumed to be in the states indicated on the top line. The shift-register stage has the characteristic of assuming the state of the preceding stage after the appearance of a shift signal. Thus, after the appearance of a shift signal, the state of the various shift-register stages appears as shown on the lower line of Fig. 1. In effect, the shift signal has moved the stored information one stage to the right.

Basic Shift-Register Composition

Before describing the detailed features and applications of the various configura-

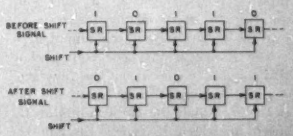


Fig. 1. Illustration of the shift-register function

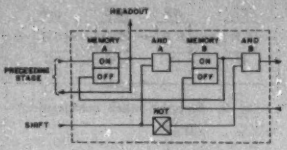


Fig. 2. Composition of shift-register stage from standard English logic functions

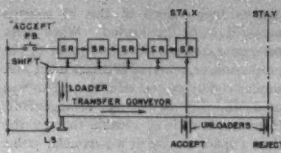


Fig. 3. Simple dispatching system involving only two destinations

tions, a word is in order regarding the basic composition of shift registers. Fig. 2 shows a group of logic elements arranged to perform the shift-register function.

Note that the basic function is composed of a NOT element, two MEMORY elements, and two AND elements. The first MEMORY denoted MEMORY A defines the state of the shift-register stage.

Assume MEMORY A is ON. Then momentary appearance of the shift signal causes AND A to have both its inputs present, and hence, to produce an output. This output turns MEMORY B ON. MEMORY B, in turn, turns MEMORY A OFF. The ON signal from MEMORY A has now been completely shifted to MEMORY B where it remains so long as the shift signal is present.

When the shift signal disappears, the NOT element puts out a signal to AND B. This causes the information in MEMORY B to be shifted to MEMORY A of the succeeding stage in a manner similar to the shifting from A to B just described.

Numerous variations of the foregoing configuration exist. For example, the output of a single NOT element may be used as the input to AND B for several stages. Thus, the NOT need not be part of each stage. In another variation, a sequence of signals appears to cause the shifting and the resetting of the MEMORIES, instead of having these functions performed internally as shown. Although some shift-register circuits are not made explicitly of the logic elements as shown, the logic functions themselves are inherent within them. Thus, the actual circuitry is often simpler than that implied by Fig. 2.

Arrangement of Shift-Register Stages for Material Dispatching

With this review of shift-register operation, it follows that information can be moved through a control system in synchronism with material being transported in a materials-handling system. For example, coded information can be introduced into a control at the time material enters a conveyor. Such information could determine where the material is to be unloaded or what processes are to be performed upon it. As the material passes along the conveyor, the information advances through the control. This is accomplished by causing the conveyor drive to deliver a shift pulse each time the conveyor has advanced a certain distance. When the material arrives at a point on the conveyor where the desired process is to be performed, the information arrives at a corresponding point within the control to permit the initiation of the process.

Fig. 3 shows a simple dispatching system involving only two destinations. Here the material is loaded onto a transfer conveyor by a loader shown at the extreme left. At the time the material is loaded, it is inspected by an operator. If the material is acceptable, the operator presses the "accept" pushbutton. If it is rejected, the operator presses no buttons. Accepted material is to be unloaded at station *X*. Rejected material is to be unloaded at station *Y*. The pressing of the accept pushbutton turned on the first shift-register stage, that is, it introduced a *1* into the system. This *1* advances through the cascaded shift-register stages as the material advances along the transfer. Note the limit switch (*LS*) which provides a shift signal for each transfer index or movement over a preset distance. When the material arrives at station *X*, the *1* arrives at the shift-register stage corresponding to station *X* and initiates the station *X* unloader. If the material was rejected, no *1* would be introduced, and a *0* would have been passed along through the control. Since the *0* will not initiate the station *X* unloader operation, rejected material passes to the end of the transfer, where it is unloaded at station *Y*. Thus, the single line of cascaded shift-register stages permits two possible choices of material destination.

Applying the binary coding principle, parallel lines of shift-register stages in cascade can be used to increase the number of choices. This is illustrated by the system shown in Fig. 4. Here 16 possible choices exist as shown by unloading stations *A* through *P*. Sixteen dispatch pushbuttons would be used to provide the 16 possible choices. Each pushbutton would introduce one of 16 possible codes into the four parallel lines of shift-register stages shown. The 16 codes are shown immediately beneath the parallel shift-register stages at each station on Fig. 4.

The arrangement shown in Fig. 4 can be somewhat simplified. Such simplification is done by causing operation of an unloader to reset (that is, erase information from) the single parallel group of shift-register stages which initiated its operation. A code of four *0*'s thus appears in a stage after its unloader has operated.

Note in Fig. 4 that the fourth line of the shift register needs only to transmit *0*'s beyond station *H*. Similarly, the third line need only transmit *0*'s beyond

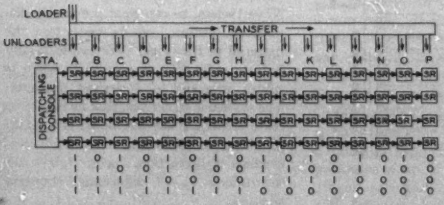


Fig. 4. Dispatching system involving 16 destinations

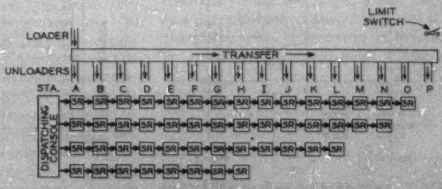


Fig. 5. Simplification of parallel shift-register arrangement

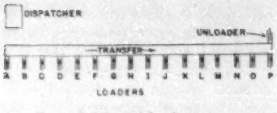


Fig. 6. Material loading system

station *L*. The second line need only transmit *0*'s beyond station *N*. And the first line need transmit only *0*'s into station *P*. Since a choice does not, in fact, exist between *1* and *0* in these stages, they may be removed. A limit switch, or other detecting device, can now be used for the initiation of operation of the unloader at station *P*. Station *P* unloader previously operated when no code existed (that is, four *0*'s were present), which condition exists after an unloader has operated and its code has been erased. Now the station *P* unloader operates only when a body appears as indicated by the limit switch. The simplified version of the shift-register arrangement is shown in Fig. 5.

Two conclusions can be made from the preceding discussion. First, the basic number of parallel shift-register lines required is equal to the number of binary digits required for representing the number of choices. Second, by erasing the code after each actual operation, this number is reduced to that required for the remaining choices only.

A complementary condition exists for a material loading scheme based upon similar premises.

Assume that a transfer conveyor exists upon which material is to be loaded as directed by an operator or another part of the control system (see Fig. 6). It is desired to load this material in some particular order from 16 possible sources. As consecutive spaces appear at the head end of the transfer conveyor, the dispatcher, be he human or mechanical, indicates from which of the 16 sources material is to be loaded. When that particular space is moved adjacent to the proper source, material is then loaded. A similar coding system can be used as was used in the unloading arrangement. However, here it is possible that by the time a space arrives at a particular point to be loaded, the supply of material may have been exhausted. Anticipation of such a condition is difficult since the backlog of orders already in the system for a given source may vary.

In this case, the additional logic elements shown in Fig. 7 are introduced. When reducing the number of parallel lines, this logic arrangement prevents the shifting of any *1*s if the last stage of the discontinued line contains a *1*. In this form of erasure it is not absolutely necessary that a scheduled operation occur in order to prevent a subsequent false operation further down the line.

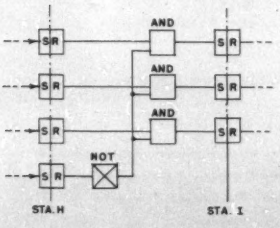


Fig. 7. Method of erasure used with loading system

Features of an Actual Field Application of Dispatching

Beside the usual high reliability of an all-static system, the shift-register dispatching means provides the ultimate in flexibility. All of the stored information is available at all times in a form compatible with the various sequential functions. Such high flexibility is very desirable in the automotive industry or in other industries where interruption of a process might result in tremendous losses. Besides demanding high reliability, high flexibility is desired in order that the operator have wide latitude in the event of mechanical breakdown. Such systems are designed to provide the highest performance possible on the remainder of the system in the event one section is disabled.

Fig. 8 shows a plan schematic of a conveyor system employing the shift-register techniques just described for dispatching. Bodies are loaded on the loading transfer by means of the loader shown at the lower left. At the time of their loading, the operator dispatches each body to one of 15 storage conveyors. This dispatching is based on body style, color of paint, and/or the type trim. Thus, the conveyor system functions both as a storage means and as a sorting means.

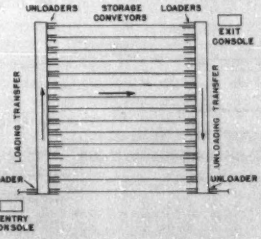


Fig. 8. Plan of actual field application

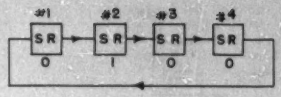


Fig. 9. Stepping-switch configuration

Bodies are unloaded from the storage conveyors onto the unloading transfer as commanded by the console at the upper right. The bodies are withdrawn in an order conducive to economical production operations. Thus, the complete conveyor system serves to sort the bodies so that economical production runs can be made. Both transfers are indexing types.

In order to prevent overloading any one storage conveyor, a differential counter is provided for each storage conveyor. This counter adds one at the time the body is loaded on the loading transfer when the loading dispatch command is given. It subtracts one when a body is loaded from the given storage conveyor to the unloading transfer. Thus, the counter maintains a net count of all bodies actually on a storage conveyor or enroute to that storage conveyor. It is so interlocked that further dispatching cannot be made if a certain maximum net count is reached.

The previous paragraphs describe the basic operation of this system. However, besides providing the automatic dispatching, the system permits the operator to change his mind and reroute material that has already been dispatched. For example, assume that for some reason no additional material is to be dispatched into storage conveyor no. 12. The operator may or may not remember the destinations of all bodies on the loading transfer between the loader and station 12. Nevertheless, he can clear previously dispatched orders, one-by-one, for the material in this zone. He can manually initiate unloading of any body when it comes adjacent to a desired unloader. The system is so flexible, that he may thus erase orders for any one body, any group of bodies, or all bodies in this zone. Regardless of his action, correct tallies are maintained on all differential counters affected.

The maintenance of correct tallies on

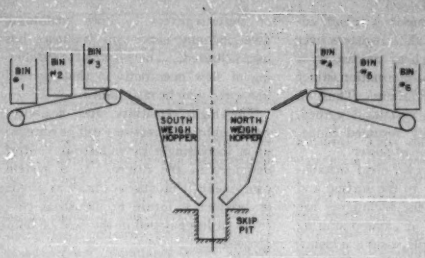


Fig. 10. Coke-feeding arrangement controlled by shift register

the differential counters is accomplished in the following fashion. Assume that a body at station 6 on the loading transfer is to have its order cleared. Further assume that the order was for the loading of the storage conveyor no. 12. The operator shifts a selector switch from AUTOMATIC to CLEAR and presses the dispatch button corresponding to the position where the body in question is now located. When the button is pressed, the code is broadcast over common signal busses to all differential counters. Subtraction is made on the differential counter for which the code applies. In this case, it would be a subtraction from the counter of storage conveyor no. 12. When the body is advanced to an alternate conveyor onto which it may be loaded, the operator sets his selector switch onto MANUAL and now the dispatch button may be pressed corresponding to the conveyor onto which the body is to be loaded. If, for example, this were conveyor no. 8, the differential counter for conveyor no. 8 would be advanced one count.

Where system flexibility is needed, as described in the preceding, shift-register techniques apply remarkably well.

Description and Field Application of the Stepping-Switch Configuration

Fig. 9 shows another important configuration. Here several shift-register stages are cascaded as previously described. Unlike the previous system however, the first stage is cascaded directly behind the last stage so that a closed ring exists. Initially, all stages are turned OFF *0* except one. This one stage is turned ON *1*. The *1*, or ON state, is thus shifted from stage to stage each time the shift signal appears. Thus, it continues to travel around the ring. It is seen that this is analogous to stepping-switch operation. Minor modifications of the preceding permit the shift register to be applied to scanning and counting functions as well.

Several field applications employing this technique can now be cited.

In one of the large eastern steel mills, this technique has been employed in an automatic blast furnace programmer. It is of general engineering knowledge that material is fed into the blast furnace by means of a skip hoist. The skip hoist is loaded with material in a pit below ground level. It is then hoisted up an incline and charges the material into the top of the furnace. A very flexible means of programming the material to be charged into this skip is obtained by using the stepping-relay analogy. Each time the skip enters the pit, the shift-register arrangement is advanced one stage. Suitable logic elements permit the control to read a loading command which indicates material to be charged on that particular load. When the skip car next returns to the pit, the shift register has advanced one stage and the command for that load is now read. It is quite possible to reset the configuration back to stage no. 1 at any point in the cycle by previous programming. Programming is accomplished by means of a battery of selector switches. In the particular application, it is possible to have from 1 to 10 skips programmed on a repeating basis as desired. In fact, two programs of up to 10 skips each may be set up simultaneously, and the system will follow either exclusively or both consecutively.

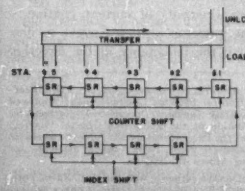


Fig. 11. Automotive conveyor using stepping-switch configuration

Shift-register techniques in the stepping mode are also applied on this same control to select one of a group of coke bins that are to be used to supply coke. Fig. 10 shows the arrangement of feeding coke. Coke is fed into the car first from the right side and then from the left side on consecutive skips. The weigh hopper on the left is refilled from one of three bins, nos. 1, 2, or 3. Normally, it is refilled from a different bin on each consecutive refill. However, it is very simple to eliminate any one bin or any group of bins from the program simply by switching the undesired bin OFF. The shift register circuit is arranged so that it will search for a bin that is ON, passing up all of those that are OFF. Thus, the system scans the bins and selects a suitable one (in this case, one that is turned ON).

Fig. 11 illustrates an application employing the stepping principle in the automotive industry. Here is shown an indexing-transfer conveyor receiving material from one of five storage conveyors. It is desired to remove ten bodies each from the various storage conveyors in consecutive order. To accomplish this, a counter which provides a signal at the count of 10 is used for advancing the shift registers. Thus, the ON signal (or *1*) advances along the upper row of shift register stages from nos. 1 to 2 to 3 to 4 to 5. At station no. 5, it is desirable next to shift back to station no. 1. However, all bodies on the transfer must be cleared. Therefore, the shift back to station no. 1 is accomplished by means of the lower row of shift-register stages. This row, however, must be shifted by means of the index transfer.

Using Shift Registers in Counting

It is obvious from what has been said previously that the shift register can be used as a basic counting element. Fig. 12 shows a basic scale-of-two, or binary, counter composed of two shift-register stages. By making the pulse to be counted the shift pulse, this circuit will shift back and forth alternately as the shift pulses are received. If a signal from one of the two stages is now used as a shift pulse for another similar circuit, this system could be used to count to four. By cascading additional pairs, a binary

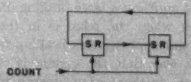


Fig. 12. Scale-of-tw

counter of any desired capacity can be composed. Quite often, in the interest of economy, it is desirable to combine this scale-of-two stage with a scale-of-three, a scale-of-four, or other combinations to provide a low-capacity counter.

Conclusions

The foregoing illustrations have presented the shift register as a performer of the following functions:

1. Information dispatching.
2. Stepping switch and scanning.
3. Counting.

High versatility has been achieved in most cases by providing alternate signals for shifting in addition to the main shift signal. Because shift-register circuits are available that are compatible with industrial-type magnetic elements, they can be incorporated directly into industrial-control schemes. A chief advantage of magnetic shift registers over some other types, such as transistors, is their ability to retain information intact even through a power failure. Thus, on any of the dispatching systems described, operation can be easily resumed immediately upon restoration of power.

Although the conveyor field installations described apply to the sorting and storage of material, the techniques are just as easily used in manufacturing processes. For example, assume a manufacturer desired to use one punching and stamping line for several different models of television chassis which he makes. The various chassis might be made by different combinations of operations in the line. By coding the blanks as they enter the line, the various operations could be indicated in much the same manner as that described for loading and unloading. Successful field application of shift registers to lines containing diverging and converging branches has been achieved. This permits the paralleling of slow operations so that over-all line speed may be maintained.

The high versatility and flexibility provided by shift registers will be counted upon increasingly by industrial control designers in the future. As a system component, it promises to have as bright or brighter a future in industrial control as it has enjoyed in the field of data processing.

References

1. Digital Computer Components and Circuits (book), R. K. Richards. D. Van Nostrand Company, Inc., Princeton, N. J., 1957.
2. Magnetic Logic Circuits for Industrial Control Systems, W. G. Evans, W. G. Hall, R. I. Van Rice. *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 166–71.
3. The Single-Core Magnetic Amplifier as a Computer Element, R. A. Ramey. *Ibid.*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 442–46.

Quasi-Optimization of Relay Servomechanisms by Use of Stored Energy for Braking

C. McDONALD
ASSOCIATE MEMBER AIEE

G. J. THALER
MEMBER AIEE

IN A PREVIOUS PAPER[1] the authors show that a relay servomechanism can be made to have a deadbeat response to a step displacement input by using proper combinations of linear derivative switching and discontinuously applied viscous damping. In brief, viscous damping is applied to the motor or load when the relay enters its dead zone. The resultant phase trajectory in the dead zone is a straight line for many cases, and derivative controlled relay operation can be adjusted to make the locus of points at which the relay opens parallel to the dead-zone trajectory. The relay cannot reverse under these conditions and the system gain can be raised until the dead zone is quite narrow.

The fundamental physical idea behind the discontinuous damping principle is that of applying a large braking force (or torque) when the relay is in the dead zone, and applying this braking force in such a way that the system is stopped without further relay operations. A similar, but possibly more effective idea, is that of using stored energy to produce a greater braking force and therefore quicker deceleration and faster over-all response. This concept may be explained by reference to the block diagram of Fig. 1. This diagram represents a normal relay servo except for the addition of an energy storage device in parallel with the motor unit. When the relay is closed the power supply drives the motor, but at the same time feeds energy into the storage unit. When the relay opens, entering the dead zone, the energy stored in the storage unit is released in such a way that it neutralizes the kinetic energy in the motor-load unit, effectively setting up a braking force and stopping the motor.

Such a scheme obviously has many ramifications. The amount of energy stored in the storage unit must at all times be related to the kinetic energy stored in the motor-load units, since too great a stored energy would probably result in motor reversal and an insufficient stored energy would not stop the motor. In addition the rate of discharge of the stored energy obviously would affect the magnitude of the braking force and thus would control the deceleration and the over-all speed of response. Finally, the equation of the system must be such that the braking may stop the motor without relay reversal, and in the presence of a relatively narrow dead zone. It is this final consideration which leads to a phase-plane approach to the problem.

A Physical System Employing the Energy Storage Brake

A very simple way to obtain energy storage braking is to use a d-c shunt motor for the relay servo, and connect a resistance-inductance circuit across the motor armature; see Fig. 2. The operation of this circuit may be explained qualitatively as follows: When the relay is closed current flows from the source through the parallel paths of armature and *R-L* circuit. When the relay opens the requirements of current continuity in the inductance cause current to flow through the inductance in the direction

Paper 58-1271, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 15, 1958; made available for printing August 29, 1958.

C. McDonald formerly with the U. S. Naval Postgraduate School, Monterey, Calif., is now a lieutenant, U. S. Navy, Kailua, Oahu, T. H., and G. J. Thaler is with the U. S. Naval Postgraduate School, Monterey, Calif.

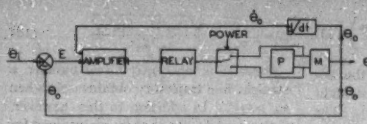


Fig. 1. Block diagram of relay servo with energy storage brake

P = energy storage device
M = motor

established by the supply, but this current then must flow through the armature in the reverse direction, setting up a braking torque. Since the magnitude of the initial value of this reverse current is determined primarily by the values of resistance and inductance in the storage circuit, the initial value of current may be made even larger than the normal motor starting current. It is readily seen that extremely large instantaneous braking torques may be set up with consequent rapid decelerations

An actual relay servo using such an inductive storage unit was set up in the laboratory. Tests established the existence of an appreciable braking force, but an exhaustive experimental study was not made. Instead the problem was analyzed on the phase plane and the results verified and extended with an analog computer study.

Phase-Plane Analysis

From physical reasoning it is apparent that the energy storage unit cannot affect the motor performance during acceleration as long as the power supply is well regulated. Acceleration trajectories on the phase plane are therefore of the well known torque-saturated type and need not be discussed. The energy storage unit acts as a brake only when the relay is in the dead zone. It is therefore necessary that a dead zone exist, and the points of interest are the geometry of the dead zone itself, and that of the phase trajectories in the dead zone.

While the boundary of the dead zone may be made to conform to a large number of shapes by proper design of a nonlinear switching computer it is a simple matter to obtain a straight-line boundary by using a derivative switching device such as a tachometer or electronic differentiator. For the purposes of this study it has been decided to restrict the dead-zone boundary to be such a straight line, but the effect of nonlinear switching devices may be estimated from the results obtained.

If the dead-zone boundary is to be a straight line, then optimum static accuracy must obtain when the phase trajectory in the dead zone is a straight line parallel to the switching line since under these conditions the width of the dead zone may be reduced to an infinitesimal amount. Thus the first problem is to determine the possibility of the existence of such a straight-line trajectory. If it is found that such a trajectory can exist then the conditions for its existence must be established and studied in the light of physical realizability. This study may be undertaken by analyzing the equation for the isoclines of the system.

To simplify the analysis it is assumed that viscous and coulomb friction are negligible, and the equilibrium equations of the motor-load unit in the dead zone are

$$(L_a+L_b)\dot{i}+(R_a+R_b)i+K_v\dot{\theta}_o=0 \tag{1}$$

$$K_t i=J\ddot{\theta}_o \tag{2}$$

where

L_a, L_b = armature and braking circuit inductance
R_a, R_b = armature and braking circuit resistances
i, $\dot{i}$ = armature current and its derivative
K_v = motor back emf constant
K_t = motor torque constant
θ_o; $\dot{\theta}_o$; $\ddot{\theta}_o$ = output position, velocity, acceleration

From these equations it can be shown (Appendix) that the slope of the phase trajectory is

$$\frac{d\dot{\theta}_o}{d\theta_o}=-\frac{R}{L}-\frac{K_vK_t}{LJ}\frac{\theta_o}{\dot{\theta}_o}+\frac{\frac{R}{L}\dot{\theta}_x+\frac{K_t}{J}i_x}{\dot{\theta}_o} \tag{3}$$

where

$R=R_a+R_b$; $L=L_a+L_b$;
$\dot{\theta}_x$ = value of $\dot{\theta}_o$ at instant dead zone is entered
i_x = value of i at instant dead zone is entered

Letting $N=d\dot{\theta}_o/d\theta_o$ a constant and defining new coefficients in terms of the natural time constants of the system, equation 3 becomes:

$$N=-\frac{1}{\tau_e'}-\frac{\theta_o}{\tau_m'\tau_e'\dot{\theta}_o}+\frac{(\dot{\theta}_x/\tau_e')+(i_x/k_m)}{\dot{\theta}_o} \tag{4}$$

where

$\tau_e'=(L_a+L_b)/(R_a+R_b)$
$\tau_m'=J(R_a+R_b)/K_vK_t$
$k_m=J/K_t$

Equation 4 contains the isocline relationships, but is not usable unless a definite relationship exists between the initial condition $\dot{\theta}_x$ and i_x. $\dot{\theta}_x$ is the speed at the instant the system enters the dead zone, and i_x is the adjusted (or resultant) braking current at the same instant. $\dot{\theta}_x$ depends on the motor-acceleration characteristics. Since the armature inductance is being included in the analysis, the motor-transfer function is a quadratic.[3] In many cases, however, the quadratic is overdamped and the motor characteristics may be described with reasonable accuracy by the mechanical time constant alone, this being $\tau_{ma}=JR_a/K_vK_t$. Temporarily restricting the analysis by this definition, it may be stated that the rise in motor speed in response to a step function is an exponential rise with time constant τ_{ma}.

Since the energy circuit is in parallel with the motor, energy is being stored in the inductance at the same time that kinetic energy is built up in the moving parts. By making the electrical time constant $L_b/R_b=\tau_{e2}$ the same as the motor time constant, i.e., $\tau_{ma}=\tau_{e2}$, the current through L_b builds up at a rate proportional to the speed build up, and the inductively stored energy builds up at a rate proportional to the kinetic energy build up. These restrictions are convenient for the mathematical manipulations.

The initial value of current i_x is not the same as the current through the inductance L_b. If the laws of conservation of momentum and conservation of energy are applied:

$$i_x(L_a+L_b)=i_bL_b-i_aL_a \tag{5}$$

$$\frac{i_x^2(L_a+L_b)}{2}+\frac{E_R}{2}=\frac{i_b^2L_b}{2}+\frac{i_a^2L_a}{2} \tag{6}$$

where

i_a = armature current prior to relay opening
i_b = current through L_b prior to relay opening
E_R = twice the energy dissipated at the switching instant. Note that some energy must be removed from the circuit at the instant the relay opens if both laws of conservation of energy and momentum are to be satisfied. This energy may be dissipated in the spark at the relay contacts, or perhaps in some other way

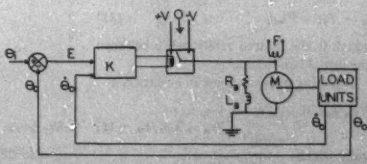


Fig. 2. D-c shunt motor relay servo with R-L circuit in parallel with motor armature

By solving equation 5 for i_x, substituting this result in equation 6 and manipulating, the amount of energy dissipated is

$$\frac{E_{\Sigma}}{2}=\frac{L_aL_b(i_b+i_a)^2}{2(L_a+L_b)} \tag{6A}$$

From equation 5:

$$i_x=\frac{L_bi_b-L_ai_a}{L_b+L_a}=Fi_b \tag{7}$$

where

$$F=\frac{L_bi_b-L_ai_a}{(L_b+L_a)i_b} \tag{8}$$

and the range of values for F is seen to be $0\leq F\leq 1$.

Further insight into the nature of F is obtained by noting that the current i_b increases from zero toward some maximum value, but the current i_a initially rises rapidly to a large value, then rapidly decreases as the motor back electromotive force builds up. For virtually all disturbances (except possibly some very small ones) $i_b>>i_a$ when the relay opens. In addition, if $L_b>>L_a$, which must necessarily be the case if L_b has been deliberately inserted for braking purposes, then for very small disturbances where the relay opens after a very short time

$$F\cong\frac{L_bi_b-L_ai_a}{L_bi_b} \tag{9}$$

and for larger disturbance for which there is a longer time from relay closure to relay drop out

$$F\cong\frac{L_bi_b}{L_bi_b}=1.0 \tag{9A}$$

In analyzing performance it is convenient to use the relationship of equation 7, i.e., $i_x=Fi_b$, assuming various values of F between 0 and 1.0. In practice the expected value of F would be very nearly 1.0, but some deviations may be expected for small disturbances.

However, the choice $\tau_{ma}=\tau_{es}$ defines

$$i_{bz}=k_e\theta_z \tag{10}$$

where K_e is a proportionality constant and is approximately K_v/R_b, and i_{bz} is the value of i_b at the instant the relay opens.

It then follows that

$$i_x=Fi_b=Fk_e\theta_z \tag{11}$$

which is the desired relationship between the initial conditions.

Inserting this condition in equation 4

$$N\tau_m'\tau_e'\dot{\theta}_0=-\tau_m'\dot{\theta}_0-\theta_0+\tau_m'\dot{\theta}_z+F\tau_m'\tau_e'k_e\dot{\theta}_z/k_m \tag{12}$$

from which

$$\dot{\theta}_0=\frac{-\theta_0}{\tau_m'(N\tau_e'+1)}+\dot{\theta}_z\left(\frac{1+\tau_e'Fk_e/k_m}{N\tau_e'+1}\right) \tag{13}$$

Equation 13 is the equation for the isoclines, and shows that they are straight lines with slope $-1/[\tau_m'(N\tau_e'+1)]$. The focal point through which all of the straight lines pass is

$$\theta_0=\dot{\theta}_z\tau_m'(1+k_eF\tau_e'/k_m);\ \dot{\theta}_0=0 \tag{14}$$

Since the isoclines are straight lines emanating from a focus it is reasonable to expect that a straight-line trajectory can exist only if it is identically one of these isoclines. This means that there must exist an isocline such that the N for that isocline is both the slope of the trajectory and the slope of the isocline. It is convenient to call this isocline the "critical isocline." Assuming that this condition can be satisfied, equation 14 specifies that the location of the focal point depends on the value of $\dot{\theta}_z$, which is the initial condition. Since the equations were derived with the implicit assumption that $\theta_0=0$ at the instant the relay opened, this is equivalent to the phase portrait of a dead zone parallel to the $\dot{\theta}_0$ axis, and equation 14 merely states that all phase trajectories in such a dead zone are parallel straight lines. With derivative-controlled relay operation the dead zone is slanted, and if the slope of the dead-zone boundary is made exactly that of the phase trajectory there is an effective translation of the origin of co-ordinates which exactly counterbalances the translation of the focal point.

This may be expressed mathematically by adding the restraint that the $\dot{\theta}_0$ intercept of the critical isocline must be identically $\dot{\theta}_z$.

To establish the condition for the existence of a straight-line trajectory

$$N=-\frac{1}{\tau_m'(\tau_e'N+1)} \tag{15}$$

from which

$$N=-\frac{1}{2\tau_e'}\pm\frac{\sqrt{(\tau_m')^2-4\tau_m'\tau_e'}}{2\tau_e'\tau_m'} \tag{16}$$

Since N is inherently a real number, a straight-line trajectory obtains only when $\tau_m'>4\tau_e'$. In addition to this, however, if the straight line is to be maintained for all magnitudes of step disturbance, the second condition must also obtain, i.e., the $\dot{\theta}_0$ intercept must be $\dot{\theta}_z$.

For this to be true

$$\frac{1+\tau_e'k_eF/k_m}{N\tau_e'+1}=1 \tag{17}$$

or

$$N=k_eF/k_m \tag{18}$$

Thus equation 16 must evaluate to a real number, and this real number must be k_e/k_mF if the phase trajectory in the dead zone is to be a straight line for all magnitudes of step input.

Thus far the analysis has postulated that a straight-line phase trajectory in the dead zone is desirable; has defined some restrictions and approximations which facilitate algebraic manipulations; and has derived three conditions which must be satisfied simultaneously to provide the desired result of a straight-line dead-zone trajectory for all magnitudes of step input. These necessary conditions are

$$\tau_m'\geq 4\tau_e' \tag{19}$$

$$N=k_eF/k_m \tag{20}$$

$$N=-(1/2\tau_e')\pm\sqrt{\tau_m'^2-4\tau_e'\tau_m'}/2\tau_e'\tau_m' \tag{21}$$

The next step is to investigate the feasibility of satisfying these conditions with a physically realizable system. The constants of a suitable shunt motor-load unit were available from a previous[1] relay servo study, and these were rounded off to provide convenient significant figures. The motor-load parameters were

$\tau_m=JR_a/K_vK_t=0.5$

$K_v=0.25$ volt-second per radian

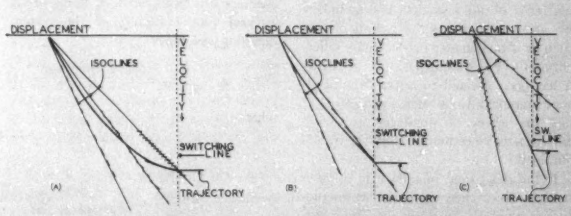


Fig. 3. Effect of parameter F on the trajectory in the dead zone

A—F=0
B—F=0.5
C—F=1.0

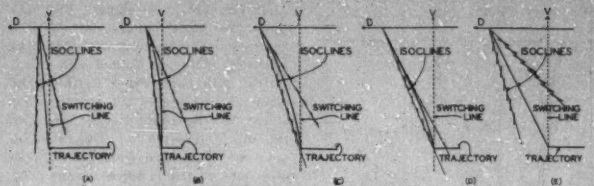


Fig. 4. Effect of parameter F on dead-zone trajectory; $L_b = 25$ henrys, $R_b = 50$ ohms, D = displacement, V = velocity

A—F = 0.975
B—F = 0.90
C—F = 0.50
D—F = 0.333
E—F = 0.167

$R_a = 300$ ohms $L_a = 0.5$ henry

$J/K_t = 0.125/300$ ampere-second^2 per radian

The parameters introduced by the analysis which are as yet undefined numerically are L_b and R_b. These quantities may be selected by the designer but not without restrictions. The restrictions are

$$L_b \gg L_a$$

$$\frac{L_b}{R_b} = \tau_{el} = \tau_{ma} = \frac{JR_a}{K_b K_t}$$

L_b and R_b may be selected to satisfy the restriction imposed on them. Certain ranges of values for L_b and R_b may then satisfy equation 19, and consequently for these values equation 21 defines N to be a real number which may be evaluated. From this value of N the corresponding value of F is computed using equation 20, and if this value lies within the specified range $0 \leq F \leq 1$, then the system is physically realizable.

Three sets of values were selected for L_b and R_b. The results are tabulated as follows:

$L_b = 150$ henrys	$L_b = 25$ henrys	$L_b = 875$ henrys
$R_b = 300$ ohms	$R_b = 50$ ohms	$R_b = 1{,}750$ ohms
$\tau_e' = 0.25$	$\tau_e' = 0.073$	$\tau_e' = 0.426$
$\tau_m' = 1.0$	$\tau_m' = 0.583$	$\tau_m' = 3.42$
$k_s/k_m = -2$	$k_s/k_m = -12$	$k_s/k_m = 0.343$
$N = -2; -2$	$N = -11.7; -2$	$N = -0.343; -2$
$F = 1.0$	$F = 0.975; 0.167$	$F = 1.0; 5.83$

Four of the five values of F indicate physical realizability. $F = 5.83$ is not possible. It is interesting to note that in the case of $L_b = 25$ henrys, $R_b = 50$ ohms, two values of F are indicated as suitable for producing the straight-line phase trajectory in the dead zone. This suggests that failure to satisfy all of the conditions stated by equations 19, 20, and 21 may not cause excessive variation from the desired straight-line phase trajectory. Note that satisfying equations 19 and 21 insures that there will be a straight-line trajectory somewhere in the dead zone if the dead zone is large enough, and the added condition of equation 20 assures that the system utilizes that straight-line trajectory for all magnitudes of step input. Therefore, if equation 20 is not satisfied because F is not as desired, the actual phase trajectory may not depart appreciably from a straight line. If a slightly curved trajectory results, the servo may be maintained as a deadbeat system by increasing the dead-zone width. An investigation of the dead-zone trajectories and isocline configurations as affected by the value of F is shown in Figs. 3(A) through 3(C) for $L_b = 150$ henrys, $R_b = 300$ ohms. Note that in Figs. 3(A) and 3(B) the changes in F have removed the critical isocline from the point at which the phase trajectory enters the dead zone. The trajectory approaches this critical isocline as an asymptote. Figs. 4(A) through 4(E) are for the case of $L_b = 25$ henrys, $R_b = 50$ ohms, and show three features very clearly.

1. The trajectory in the dead zone rapidly becomes asymptotic to the critical isocline.
2. This critical isocline approaches the point at which the phase trajectory initiates as F approaches its critical value. (The critical value of F is that value defined by equation 20.)
3. As F is reduced the average magnitude of the slope of the trajectory decreases, indicating that a longer time interval is required to stop the motor.

The third case, that of $L_b = 875$ henrys and $R_b = 1{,}750$ ohms, is not shown pictorially. The general results are the same as for Fig. 4, the difference lying in the facts that the response is much slower and the trajectories for $F < 1$ are concave downward instead of concave upward.

These results prove that the straight-

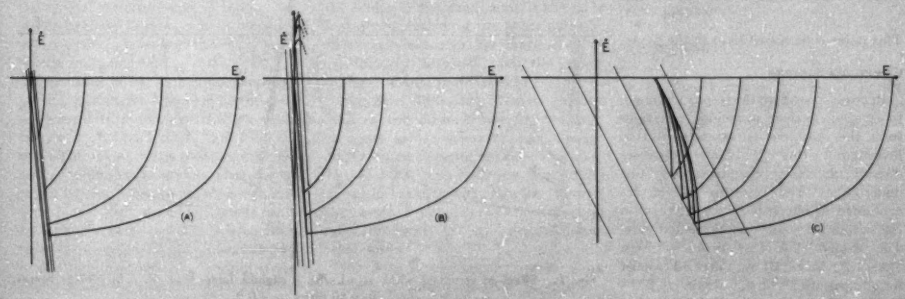


Fig. 5. Effect of large variations in $\tau_{el} = L_b/R_b$

A—$L_b = 25$ henrys, $R_b = 50$ ohms
B—$L_b = 50$ henrys, $R_b = 50$ ohms
C—$L_b = 12.5$ henrys, $R_b = 50$ ohms

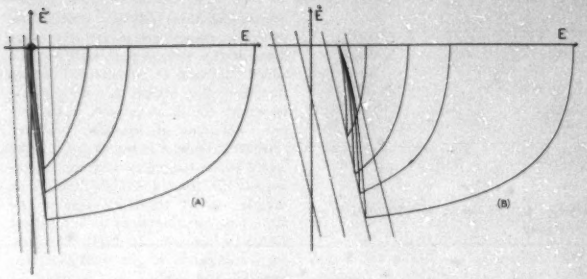


Fig. 6. Effect of small variations in $\tau_{e2}=L_b/R_b$

A—$L_b=27.5$ henrys, $R_b=50$ ohms
B—$L_b=22.5$ henrys, $R_b=50$ ohms

line trajectory in the dead zone is realizable with existing hardware, and indicate that variations in F from the optimum value may not be too harmful.

In retrospect, for $\tau_m' \geq 4\tau_e'$, if

$$\tau_{e2}=\tau_m=L_b/R_b,\ \tau_{e1}'=L_a/R_a,\ X=R_b/R_a$$

it can be shown that

$$N=FK_e/K_m=-F/\tau_m X \tag{22}$$

and

$$N=-\frac{1+X}{2\tau_m X}\pm\frac{\sqrt{\tau_m^2(1+X)^2-4\tau_m X}}{2\tau_m^2 X} \tag{23}$$

From equation 23

$$N=-1/\tau_m \text{ or } -1/\tau_m X \tag{24}$$

And it is further seen that $R_b/R_a=X$ is a very significant parameter. Analyzing equations 22 and 24 it becomes evident that a straight line is obtained for a dead-zone phase trajectory under the following conditions:

When $F=1$. The slope is then $-1/\tau_m X$
When $F=X(0\leq X\leq 1)$. The slope is then $-1/\tau_m$ (25)

This point is discussed later in the paper.

Computer Studies

In order to confirm the results obtained, to extrapolate them to include variations from the optimum parameters, and to investigate the transient responses directly, an analog computer study was undertaken. The resulting curves are contained in reference 3. A selection of the significant results are presented here. The straight line trajectories of Figs. 3(A), 4(A), and 4(E) were verified to exist for all magnitudes of step input.

When the L_b/R_b ratio varies considerably from the optimum, i.e., $\tau_e'\neq\tau_m$, the phase trajectories depart from the desired straight line characteristic. Figs. 5(A) through 5(C) show a typical variation. Fig. 5(A) shows the optimum case, and the dead-zone trajectory is straight. Fig. 5(B) shows the case of too much inductance, and it is seen that the deceleration is extremely rapid, but the system overshoots. Fig. 5(C) shows the case of too little inductance, and it is seen that the initial acceleration is rapid, but the trajectory then reverts to a gentler slope. It is easy to conclude that with too much inductance a somewhat oscillatory system must result, and with too little inductance the system not only becomes sluggish but the trajectories spread out, requiring a wider dead zone if deadbeat operation is wanted.

It was felt that some consideration should be given to smaller variations in L_b/R_b and to simultaneous variations of F. Figs. 6(A) and 6(B) may be compared with Fig. 5(A) to see the effect of varying L_b(R_b constant at 50 ohms). Fig. 6(A) shows the phase portrait for $L_b=27.5$ henrys and $F=1.0$, and it is evident that this 10% increase in L_b had little effect. On the other hand, Fig. 6(B) shows the phase portrait for $L_b=22.5$ henrys, and it is seen that the trajectories are spread by the 10% decreases in L_b. This would require an appreciable difference between drop-out and pull-in levels for the relay if chattering is to be avoided, but the trajectories all recombine before reaching zero velocity, so the static error is not necessarily affected.

Variation of F had similar effects for all of the L_b/R_b ratios tested. These effects may be summarized as follows: If the parameter relationships are such that the restrictions imposed can be satisfied by two values of F in the range $0\leq F\leq 1$, then for each of these values the trajectories are straight lines for all magnitudes of step input. The straight line corresponding to the larger value of F has the steepest slope, thus promising faster deceleration. When F is varied from 1 to 0, for those values of F greater than the larger straight-line value the trajectories have a spread and are concave toward the θ_0 axis. For values of F between the two "straight line" values the trajectories have a spread and are convex toward the θ_0 axis. For values of F smaller than the lower straight line value there is again a spread in the trajectories and they are concave toward the θ_0 axis.

A very interesting and important result is obtained by comparing Figs. 6(A), 7(A) and 7(B). Each of these is for the case of too much inductance, $L_b=27.5$ henrys. The values of F are 1.0, 0.9 and 0.75 respectively. Fig. 6(A) shows the effect of too much L_b when F remains at its optimum value; the trajectories

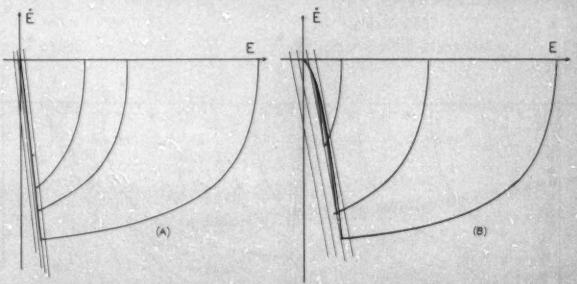
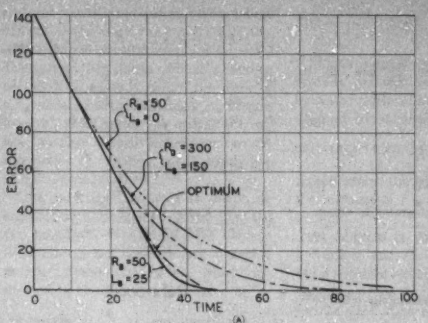
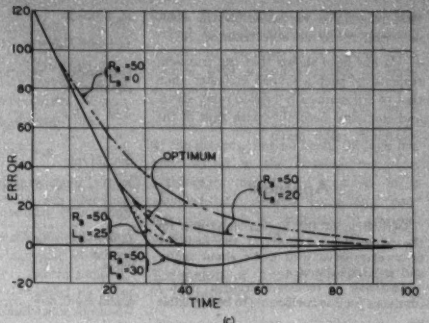


Fig. 7. Effect of varying F when $\tau_{e2}=L_b/R_b$ is slightly larger than τ_m. $L_b=27.5$ henrys, $R_b=50$ ohms, $\tau_m=0.5$

A—$F=0.9$
B—$F=0.75$





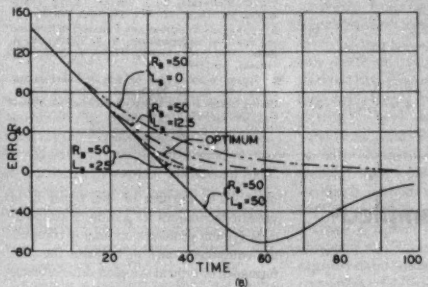


Fig. 8. Transient response of relay servo with energy-storage braking

A—Effect of varying L_b and R_b while keeping L_b/R_b constant at the best value. Comparison curves added

B—Effect of large departures from the optimum ratio of L_b/R_b

C—Effect of small departures from the optimum ratio of L_b/R_b

have a slight spread, and a slight overshoot. Fig. 7(A) shows the effect of a slight reduction in F when L_b is too large. In this case L_b is 10% too large, and F has been reduced 10% to $F=0.9$. Note that the resultant trajectories are almost perfectly straight, and are just as good as for the case of perfect conditions which is shown in Fig. 5(A). Fig. 7(B) shows that further reductions in F bring back the typical curvature and spreading. It should be noted that no such phenomenon occurs when L_b is reduced so that τ_e is below its optimum value.

The result of Fig. 7 was not predicted by the mathematics because of the difficulty involved in manipulating the equations when the parameters vary from the restricted values. The physical explanation of the result is obvious once the result is known: with too much inductance there is too much stored energy for deadbeat performance, but if this is counterbalanced by an additional loss, implied by a reduction in F, then the residual stored energy is just the ideal amount. Furthermore, if the L_b/R_b ratio is not greatly different from the optimum value the stored energy is fed back into the motor at approximately the ideal rate, and the phase trajectories are very nearly the optimum straight line.

The transient response of the system, as a function of the numerous variables, is presented in Figs. 8(A) through 8(C). These figures detail response speed only, the possibility of a static error being ignored. In practice the magnitude of the static error would be set by controlling the width of the dead zone, and this is best accomplished on the phase portrait. If it is assumed that the static error is satisfactorily adjusted, then the curves of Fig. 8 are quite significant.

Fig. 8(A) shows the effect of varying both L_b and R_b while keeping the ratio constant. It is evident that $L_b=25$, $R_b=50$ gives a much faster response than $L_b=150$, $R_b=300$, and both are faster than dynamic braking with $R_b=50$. In fact the case of $L_b=25$, $R_b=50$ compares favorably with the so-called optimum relay servo.

Fig. 8(B) shows the effect of variation in the relay parameter F for the optimum values $L_b=25$ henrys, $R_b=50$ ohms. All cases are deadbeat, but decreasing F makes the system slower. Note, however, that for all of the values of F used the response is faster than that with pure dynamic braking of $R_b=50$ ohms. For all of these cases the response seems reasonably fast so that variation in the values of F may not be important providing the required dead-zone width is not objectionable.

Fig. 8(C) shows the effect of varying the value of L_b through a range from 80–120% of the optimum value of 25 henrys, while keeping R_b constant at 50 ohms and F constant at 1.0. It is seen that all responses in this range are quite good; better than that of pure dynamic braking, and comparable to that of the "optimum relay servo." When L_b is greater than the ideal value a slight overshoot is obtained. Note that this overshoot and return to approximate correspondence all occur in the dead zone. The relay does not reclose.

Conclusions

1. Stored energy braking of a relay servo has been demonstrated to be theoretically and practically possible.

2. By applying certain practical restrictions it has been shown that such braking can be adjusted to give straight-line phase trajectories in the dead zone. This permits the use of linear derivative switching plus a narrow dead zone, with accompanying simplicity of equipment and accurate performance.

3. It has been shown that reasonable variations in the parameters of the braking circuit are not objectionable, and also

that reasonable variations in the characteristic F are not objectionable.

4. It has been shown that the transient response of such a system is faster than that of a dynamically braked relay servo, and compares favorably with that of the "optimum relay servo."

Appendix

Derivation of the isocline equations for the phase-plane analysis of the behavior of a relay servo with a d-c shunt motor and inductive braking.

Assuming C (coulomb friction) and f (dynamic friction coefficient) to be negligible

$$(L_a+L_b)\dot{i}+i(R_a+R_b)+K_s\dot{\theta}_b \quad (26)$$

$$K_t i=J\ddot{\theta}_b \quad (27)$$

Differentiating equation 27

$$K_t\dot{i}=J\dddot{\theta}_b \quad (28)$$

Let

$$L_a+L_b=L \quad (29)$$

$$R_a+R_b=R$$

$$L\dot{i}+iR+K_s\dot{\theta}_b=0$$

Substituting equations 27 and 28 in equation 29

$$\frac{LJ\dddot{\theta}_b}{K_t}+\frac{RJ}{K_t}\ddot{\theta}_b+K_s\dot{\theta}_b=0 \quad (30)$$

Simplifying

$$LJ\dddot{\theta}_b+RJ\ddot{\theta}_b+K_sK_t\dot{\theta}_b=0 \quad (31)$$

Equation 31 may be integrated to obtain

$$LJ\ddot{\theta}_b+RJ\dot{\theta}_b+K_sK_t\theta_b=C_1 \quad (32)$$

when

$\ddot{\theta}_b$ and $\dot{\theta}_b=0$, $\theta_b=\dfrac{C_1}{K_sK_t}$ assuming $\theta_b=0$ at $t=0$

From the analytical solution[2]

$$\theta_{ss1}\,(t=\infty)=\frac{RJ\dot{\theta}_z+LK_t\dot{i}_z}{K_sK_t} \quad (33)$$

Therefore

$$C_1=RJ\dot{\theta}_b+LK_t\dot{i}_b=LJ\left(\frac{R}{L}\dot{\theta}_z+\frac{K_t}{J}\dot{i}_z\right) \quad (34)$$

Equation 32 may now be written

$$\ddot{\theta}_b=-\frac{RJ}{LJ}\dot{\theta}_b-\frac{K_sK_t}{LJ}\theta_b+\frac{R}{L}\dot{\theta}_z+\frac{K_t}{J}\dot{i}_z \quad (35)$$

$$\frac{d^2\theta_b}{d\theta_b^2}=\frac{d\dot{\theta}_b}{d\theta_b}=-\frac{R}{L}-\frac{K_sK_t}{LJ}\frac{\theta_b}{\dot{\theta}_b}+\frac{\dfrac{R}{L}\dot{\theta}_z+\dfrac{K_t}{J}\dot{i}_z}{\dot{\theta}_b}$$

Now let $N=d\dot{\theta}_b/d\theta_b$ (36)

$$N=-\frac{R}{L}-\frac{K_sK_t}{RJ}\left(\frac{R}{L}\right)\frac{\theta_b}{\dot{\theta}_b}+\frac{\dfrac{R}{L}\dot{\theta}_z+\dfrac{K_t}{J}\dot{i}_z}{\dot{\theta}_b} \quad (37)$$

References

1. Quasi-Optimization of Relay Servos by Use of Discontinuous Damping, W. L. Harris, Jr., C. McDonald, G. J. Thaler. *AIEE Transactions*, vol. 76, pt. II, Nov. 1957, pp. 292–96.

2. Transfer Function and Parameter Evaluation for D-C Servomotors, G. J. Thaler, W. A. Stein. *Ibid.*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 410–17.

3. Quasi-Optimization of Relay Servomechanisms, C. A. K. McDonald. *Master's Thesis*, United States Naval Postgraduate School, Monterey, Calif., 1957.

Analog Computer Study of Sampled-Data Systems

HAROLD CHESTNUT, MEMBER AIEE — AMADEO DABUL, NONMEMBER AIEE — DAVID LEIBY, ASSOCIATE MEMBER AIEE

TO OBTAIN a satisfactory solution to many of the complex control system problems which are currently being encountered, it has been necessary to apply the combination of digital controls and analog controls to the system. Whenever a transition is made between a digital system and an analog system, the embodiment of a sampled-data system is obtained. Likewise, in applications in which a number of similar processes are controlled through a time-sharing or multiplexed system by one common computer or controller, a sampled-data system is also obtained. In such cases the system performance is determined not only by the characteristics of the continuous portion of the system within the control loop, but also by the sampled-data system parameters, such as type of holding circuit and sampling rate. To set up specifications properly for such control systems, it is necessary to understand the effect of the sampled-data system parameters upon the over-all system performance.

During the past several years many papers have been written and presented on the general subject of sampled-data systems and their analysis. The great majority of these papers have been based on the development and applications of methods of analysis using frequency response methods, z-transform techniques, or their modifications.[1] The mathematical model of the sampled-data system constructed through the use of these various analytical methods and techniques gives very useful information in the prediction of the system performance within the limitations of the assumptions made during the construction of the mathematical model.

The advent of the analog computer in the past decade has provided a powerful tool for the control engineer in the analysis of complex control systems problems. Since many engineers are familiar with analog simulation techniques, it would seem logical that the analysis of systems containing sampled-data components should also be analyzed on similar equipment. The purpose of this paper is to present techniques for handling sampled-data system problems on analog computing equipment, and to indicate the performance of typical sampled-data systems for different values of system parameters. The results obtained from the analog computer method of analysis are correlated with the results obtained by the application of different classical methods of analysis.

Review of Analytical Methods

A sampled-data system, or sampling servomechanism, can be defined as a feedback control system in which the data consist of a train of pulses or sequence of numbers at some point in the system. Such systems generally include a sampling switch which samples the continuous signal $e(t)$ at regular intervals spaced T seconds apart and gives an output signal $e^*(t)$, consisting of a train of pulses with amplitudes equal to the value of $e(t)$ at the sampling instants.

If the sampler is inserted in the error signal path of a feedback control system,

Paper 58-1324, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 7, 1958; made available for printing September 4, 1958.

Harold Chestnut, Amadeo Dabul, and David Leiby are with the General Electric Company, Schenectady, N. Y.

an error signal is applied at the sampling instant to the continuous portion of the system, which has a transfer function $G(s)$. The system then passes through a waiting period of T seconds before the error is again determined. This delay, or waiting interval, complicates the stability problem, so that special techniques become necessary to analyze and synthesize sampled-data systems.[2]

The two basic methods of analyzing sampled-data systems are based upon the two possible forms taken by the Laplace transform of the action of the sampling switch. These two forms, which are derived in the literature,[1-4] are

$$E^*(s) = \sum_{n=0}^{\infty} e(nT)\, e^{-nsT} \tag{1}$$

$$E^*(s) = \frac{1}{T} \sum_{n=-\infty}^{\infty} E(s + jn\omega_s) \tag{2}$$

The first form is convenient for manipulation in the time domain, and is the basis for the z-transform theory, proposed by Hurewicz and later expanded by Ragazzini and Zadeh.[3] The second form gives the theoretical foundation for the extension of conventional frequency analysis methods of continuous feedback systems to sampled-data systems, as pioneered by Linvill.[4]

A similar graphical method for sampled-data system analysis was developed by Tsypkin.[5]

Inspection of equation 1 shows that the relation between $E^*(s)$ and $E(s)$ is not algebraic, but transcendental, involving functions of e^{sT}. By introducing the transformation, $z = e^{sT}$, it is possible to express the transfer function of the sampler as a rational function of z. When this is done, the z-transform of the signal $e^*(t)$ is obtained:

$$E(z) = E^*(z) = E^*\left(\frac{1}{T} \ln z\right) = \sum_{n=0}^{\infty} e(nT) z^{-n} \tag{3}$$

The z-transform of the output of a network is obtained by multiplication of the z-transform of the input signal and the z-transform function of the network, $G(z)$. The usual way to find $G(z)$ is to expand $G(s)$ in partial fractions and to use a table of z-transforms. Since $G(z)$ is a function of the sampling period T and the operator s, the Nyquist plot of the $G(z)$ of the open loop of the system can be constructed and stability information can be obtained.

For a feedback control system having unity feedback, the sampled output is given by:

$$C(z) = \frac{G(z)}{1 + G(z)} R(z) \tag{4}$$

Substitution of the polynomials in z into $G(z)$ and $R(z)$ and performance of the indicated algebraic manipulation will produce an infinite series of the form:

$$G(z) = A_1 z^{-1} + A_2 z^{-2} + \ldots + A_n z^{-n} + \ldots \tag{5}$$

In this series, the coefficient of z^{-n} gives the response of $c(t)$ at the nth sampling instant.

In addition to the Nyquist stability indication and the transient response being obtained by the z-transform method, techniques have been developed by Jury for the adaptation of root-locus methods of system analysis to the z-transform concept.[6] In its elementary form, the z-transform method of obtaining the transient response gives the controlled output only at the sampling instant. However, methods have been developed by Lago for the modification of the z-transform method to give values of the control variable at points between the sampling instants.[7] Since the mathematical manipulation becomes quite cumbersome when the linear portions of the system are of high order and the driving functions are complex, programs have been written for obtaining the transient performance of sampled-data systems utilizing digital computers.[8]

In the extension to pulsed systems of the frequency domain technique used in handling continuous systems, all signals, whether sampled or continuous, are described by continuous functions and their Laplace transforms. To apply the Nyquist stability criteria to a sampled-data system by Linvill's method, a graphical approximation is obtained for the values of the function $HG^*(s)$, or $G^*(s)$ in the case of a unity feedback system. From the form of equation 2, the analytical expression for the later case becomes:

$$G^*(j\omega) \cong \frac{1}{T} [G(j\omega) + G(j\omega - j\omega_s)] \tag{6}$$

From the $G(j\omega)$ function, the normal Nyquist plot of the continuous system is obtained. The second term provides a correction vector which must be added to the curve at each given value of $j\omega$ to form the approximate curve of $G^*(j\omega)$. The usual stability checks can be applied to this curve and the closed-loop frequency response curve can be constructed.

In the methods of analysis of sampled-data systems reviewed heretofore, the sampling was considered to be instantaneous, that is, the pulse could be described by the product of a constant and a unit impulse. In a recently published paper, Farmanfarma reports a method whereby a pulse transform technique is applied to the analysis of sampled-data system with sampling pulses of finite duration.[9] This p-transform technique is manipulated somewhat in the same manner as the z-transform. However, at the present time no break-through has been found in the development of a satisfactory method of analysis of a sampled-data system containing nonlinearities such as saturation, hysteresis, or deadband.

Extension of Analog Simulation Techniques

To evaluate the performance of a sampled-data system on an electronic analog computer, it is necessary to simulate the various system characteristics peculiar to the sampled-data system, as well as those elements of the system peculiar to the continuous portion of the system. The two new elements which must be simulated are the sampling switch and the holding circuit. In the work which is described here the sampling function is obtained through the operation of a high-speed mercury-wetted relay which closes for a very short period of time at the sampling frequency. A zero-order hold function is obtained by one operational amplifier from the analog computer interconnected with the contacts of the sampling relay. The amplifier is connected as shown in Fig. 1 to operate as a unity gain amplifier with a very small time constant during the time that the sampling switch is closed, and to operate as an open input integrator when the sampling switch is open. The waveshape of the voltage input, the sampling pulse, and the voltage output are shown in the three respective figures below the computer diagram of the holding amplifier.

With a problem so scaled that the sampling rate is less than 5 cycles per second, a 10-millisecond sampling pulse will require less than 5% of the sampling period. With the values of feedback resistance and feedback capacitor as

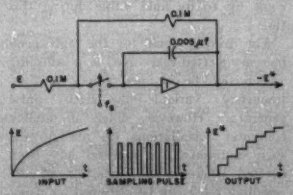


Fig. 1. Zero-order hold simulation

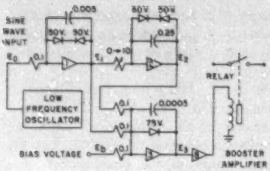


Fig. 2. Pulse generator simulation

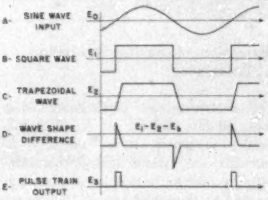
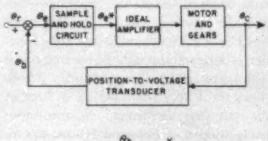


Fig. 3. Pulse generator waveshapes



$$G(s) = \frac{\theta_b}{\theta_e^*} = \frac{K}{S(1+Ts)}$$

Fig. 4. Ideal second-order system

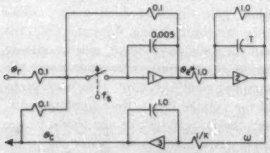


Fig. 5. Second-order system simulation

shown in Fig. 1, the time constant of the amplifier with the sampling switch closed is 1/2 millisecond. Since the sampling switch is closed for ten time constants, the amplifier output can be assumed to come up to its final value during this time.

The driving signal for the high-speed mercury-wetted relay used as a sampling switch could be obtained from a variable-frequency variable-pulse-width pulse generator. However, since all analog computer installations do not possess such equipment, a method has been devised whereby the operation of the sampling relay can be obtained by using several operational amplifiers driven from a variable-frequency sine-wave generator. The sine wave is converted into a train of pulses of one polarity and variable pulse width through the use of three operational amplifiers and one booster amplifier as shown in Fig. 2. The sinusoidal output of the variable-frequency oscillator is fed into amplifier *1*, which acts as a switching amplifier, converting the sine wave into a square wave. This square wave, which is illustrated in Fig. 3(B), is limited in magnitude to ±50 volts by the two silicon diodes connected back-to-back across the amplifier. The small capacitor is necessary to eliminate unwanted noise from the action of the limiting diodes. The square wave of voltage is applied to amplifier *2* which acts as a limited integrator with relatively high gain, to transform the square wave of voltage into a trapezoidal voltage wave. The slope of the sides of the trapezoidal wave, which is illustrated in Fig. 3(C), is a function of the feedback capacitance around amplifier *2* and of the magnitude of the input resistance. The trapezoidal voltage wave, the square wave, and a constant bias voltage are combined with the proper signs and proper gains in switching amplifier *3*. If this amplifier were operated strictly as a summing amplifier, the output voltage would be that as shown in Fig. 3(D). The subtraction of the square wave and the trapezoidal wave produces a train of triangular pulses which are alternately positive and negative. The introduction of the bias voltage shifts the origin such that the net voltage input to amplifier *3* becomes positive only during a portion of the positive triangular pulse. The action of the switching amplifier with the limiting diode connected across it causes the output of amplifier 3 to go to the full limiting voltage whenever the net input voltage is positive, and to zero voltage during all other times of the input cycle. This produces a train of rectangular pulses whose spacing is equal to one complete cycle of the input sine wave, and whose pulse length is determined by the setting of the input resistor to amplifier *2* and the magnitude of the bias on amplifier *3*. The output of amplifier *3* is then applied to a booster amplifier which supplies the necessary power to the operating coil of the switching relay. The conduction characteristic through the relay contacts has the same characteristic as the pulse train shown in Fig. 3(E).

Analysis of Second-Order System

To illustrate the analysis of a sampled-data system on the analog computer, the method will first be applied to a simple second-order system. This simple system will also allow easy comparison of the analog computer results with those results obtained from the mathematical model representation of the system. The second-order system can be represented in a practical manner by the simple position servo shown in Fig. 4. This simple position servo consists of an ideal amplifier, a driving motor and gear train, and a position-to-voltage transducer for the feedback element. The sample and zero-order hold circuit is inserted in the error signal path ahead of the amplifier. The transfer function of the continuous portion of the system consists of a gain factor, an integrator, and a single time-constant term. To investigate the effect of the sampled-data system parameters upon this second-order system, the magnitude of the gain and the time constant are adjusted to obtain different values of continuous system phase margin at crossover, while maintaining the crossover frequency constant at one radian per second. The analog computer simulation of this simple second-order system is shown in Fig. 5. Amplifier *1* performs the zero-order hold function which was discussed previously. Amplifier *2* contains the time-constant portion of the transfer function, whereas amplifier *3* simulates the integrator portion of the transfer function. The value of T in microfarads determines the time constant of the transfer function, and the value of $1/K$ in megohms determines the gain.

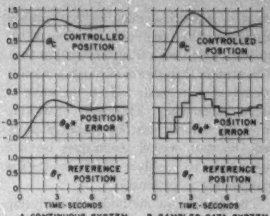


Fig. 6. Second-order system response

The response of the system to step inputs of reference position is investigated for different values of linear system phase margin and sampling-frequency to crossover-frequency ratio. Fig. 6 shows two typical responses of the system. In Fig. 6(A) the controlled position and the position error are shown for the continuous system with a phase margin at 45 degrees. In Fig. 6(B) the response of the controlled position and the sampled and held posi-

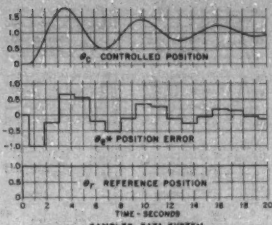


Fig. 7. Second-order system response where the sampling-frequency to crossover-frequency has been reduced to a value of 5

tion error is shown for the same phase margin, but for a sampled-data system whose sampling-frequency to crossover-frequency has a ratio of 10. A comparison of these two figures shows that the transient overshoot of the linear system is only slightly greater than 20%, whereas the sampled-data system has a transient overshoot of over 40%. In Fig. 7, the similar response is shown for the same system in which the sampling-frequency to crossover-frequency has been reduced to a value of 5. In this case, the system has deteriorated considerably in stability, becoming quite oscillatory. In this case the maximum positive transient overshoot is nearly 75%.

From a large number of transient responses taken for different values of sampling-frequency to crossover-frequency ratio and for six different values of linear system phase margin, a family of curves of system performance indication can be plotted as shown in Fig. 8. The dimensionless sampling frequency is plotted as a reciprocal curve along the abscissa, while the per-cent positive peak overshoot is plotted linearly along the ordinate. The termination points indicated by the circles are the points at which unstable operation of the system is obtained. With the system a pure integrator, no overshoot is obtained until the dimensionless sampling frequency is reduced to approximately 6. The per-cent positive peak overshoot then increases in a linear manner when the sampling frequency is plotted in this reciprocal scale. As soon as the peak overshoot exceeds the 100% value, the system became unstable with the oscillations diverging. As the system phase margin is reduced by the insertion of more phase angle into the time-constant portion of the transfer function, the system becomes more oscillatory, with greater positive overshoots, as would be expected.

Several unusual features of this family of curves should be noted. First, the curves of per-cent positive peak overshoot for phase margins of 45 degrees, 60 degrees, and 75 degrees asymptotically approach the line for the 90-degree phase margin through a peculiar flattening of the curve. Second, it should be noticed that it is possible to have a sampled-data system with less than a 90-degree phase margin operate stably at a much lower value of sampling frequency ratio than the pure integrator.

This family of curves can be quite useful as a design tool in predicting the effect of sampling frequency upon the performance of a system which can be assumed to behave as a second-order system having a given value of phase margin. For example, a linear second-order system with a phase margin of 45 degrees has the positive peak overshoot increased from approximately 23% to 46% if it is operated as a sampled-data system with a sampling-frequency to crossover-frequency ratio of 10.

This family of curves can also be used to interpret the data in a somewhat different manner. A sampled-data system with a dimensionless sampling frequency of 10 and a phase margin of 45 degrees has the same transient performance, as far as positive overshoot is concerned, as a continuous system whose phase margin is somewhat below 30 degrees. The family of curves can also be used to give information on the required improvement of continuous system phase margin to obtain a given sampled-data system with the same performance as the original continuous system. For example, a 45-degree phase margin second-order system would have to be improved by somewhat over 15 degrees' phase margin in order to obtain the same performance when operating as a sampled-data system with a sampling frequency ratio of 10.

To check the validity of the analog computer simulation of the sampled-data system and the results obtained therefrom, a comparison is made with the results obtained by several analytical methods. In Fig. 9, the Nyquist plot is drawn of the second-order system with a phase margin of 45 degrees for different values of sampling frequency to crossover frequency and also for the continuous system. These data were obtained by using Linvill's method and also from the z-transform method of analysis. This family of curves shows that the -1 point is encircled somewhere between the sampling ratios of 2.5 and 3.3, being very close to the 3.3 value. In a detailed

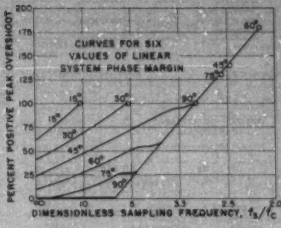

CURVES FOR SIX VALUES OF LINEAR SYSTEM PHASE MARGIN
PERCENT POSITIVE PEAK OVERSHOOT
DIMENSIONLESS SAMPLING FREQUENCY, f_s/f_c

Fig. 8. Effect of sampling frequency on second-order system performance

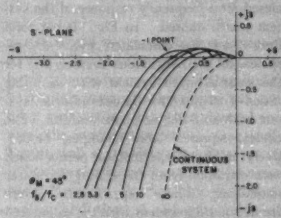


Fig. 9. Nyquist representation

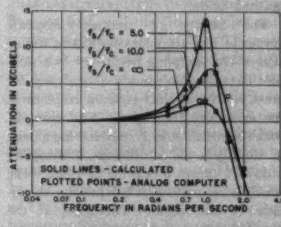
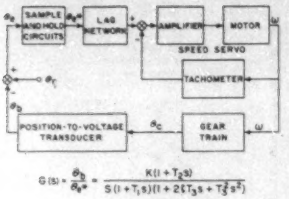


Fig. 10. Frequency response curves

analysis, it was found that the exact sampling ratio at which instability occurred was approximately 2.95. From the results of the analog computer study, it was found that the system did not really become unstable and diverge until the sampling-frequency ratio was reduced to nearly 2.5. However, in the region between approximately 3.0 and 2.5, the system was extremely oscillatory with no clear indication as to whether the oscillation was to diverge or to converge with semiresonant increasing and decreasing magnitude of the oscillation as a function of time. From this comparison, it would appear that the analog computer simulation shows a tendency for a slightly lower possible sampling frequency before unstable operation is obtained.



$$G(s) = \frac{\theta_c}{\theta_e^*} = \frac{K(1+T_2 s)}{s(1+T_1 s)(1+2\xi T_3 s + T_3^2 s^2)}$$

Fig. 11. Position control system

In a second comparison of the calculated and analog computer results, the closed-loop frequency response of the system was obtained. In Fig. 10 the solid lines show the calculated frequency response of the continuous system, and for the sampled-data system with sampling ratios of 10 and 5 for the second-order system with a 45-degree phase margin. The circles, squares, and triangles, respectively, indicate experimental points which were obtained from the analog computer simulation. These experimental points are in good agreement with the calculated results, showing that in the frequency domain the analog computer simulation behaves in a manner similar to the mathematical model.

The family of frequency-response curves in Fig. 10 points out that a reduction of the sampling frequency of a sampled-data system causes a tendency for great peaking of the frequency response at the crossover frequency. For example, with the continuous system having a peak of slightly over 2 decibels, a sampling ratio of 5 causes a peaking of 14 decibels, which is an increase of 5 times the magnitude of the input. Therefore, if the sampled-data system is used in an application where the noise spectrum includes the crossover frequency, there will be considerable noise amplification as a sampled-data system, whereas as a continuous system, the noise will not be greatly amplified.

A third area in which the performance of the analog computer simulation was compared with the results obtained analytically was in the determination of the transient response. The z-transform method was used to calculate the magnitude of the controlled variable at the first 25 sampling periods following the application of the step driving function to the input reference position. These points were then plotted on the recorder chart obtained from the analog computer representation of the same system conditions. When the sampling frequency calibration was maintained to a good degree of accuracy, it was found that the analytical results and the observed results checked exceedingly well, the errors being less than the reading errors of the 5-centimeter chart paper.

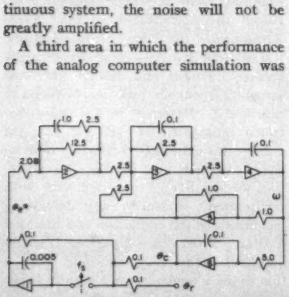


Fig. 12. Position control simulation

Application to Higher-Order System

The analog computer simulation method of analyzing sampled-data systems will now be extended to a system containing higher-order transfer functions terms, as well as certain nonlinearities. The system to be considered is a typical position control such as could be used in machine tool applications. An elementary block diagram and the transfer function of such a system are shown in Fig. 11. The error between the reference position and the position feedback is applied to a sample-and-hold circuit. The signal then goes through a lag-compensating network into a speed servo loop, which consists of an amplifier, a motor, and a tachometer. The output of this speed loop drives the output position through a gear train. The open-loop transfer function shown at the bottom of the figure consists of a gain term, the lead term from the compensating network, an integrator, the lag term from the compensating network, and a pair of complex roots from the speed servo. This system is simulated on the analog computer with a 10-to-1 reduction in time scale in order to operate the sampling relays at a satisfactory rate, and also to be able to obtain the transient response satisfactorily on the recording equipment. The simulation on the analog computer is shown in Fig. 12. Amplifier *1* gives the sample-and-hold function, amplifier *2* contains the stabilizing network, amplifiers *3*, *4*, and *5* simulate the speed servo loop, and amplifier *6* simulates the integration of the speed to give the output position.

In the previous example of the second-order system, the step function alone was used as the driving input. However, in many position controls, this is not a realistic type of driving function, particularly in those cases in which the position control is part of a machine tool system. It is important, therefore, to determine the behavior of the sampled-data system for various types of input. In Fig. 13, the response of the controlled position, the motor speed, and the position error are given for a step change in the reference position. In Fig. 13(A), the response is given for the continuous system. In Fig. 13(B), the response is given for the sampled-data system with a sampling-frequency to system-crossover-frequency ratio of 10. The continuous system itself is slightly underdamped, and as such, does not look like a very satisfactory position control when a step function is the input. When it is operated as a sampled-data system with the sampling ratio of 10, the transient overshoot is considerably greater and the oscillation is less damped.

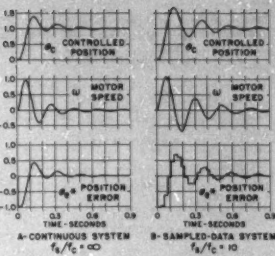


Fig. 13. Position control response to step input

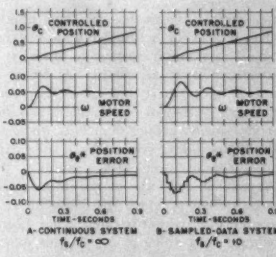


Fig. 14. Position control response to constant velocity input

In Fig. 14, the responses of the continuous system and the sampled-data system are shown for operation with a constant velocity input. In this case, it should be noted that the scales of the motor speed and position error indication have been expanded by a factor of 10 compared to the previous figure in which the input was

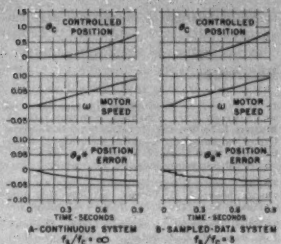


Fig. 15. Position control response to constant acceleration input

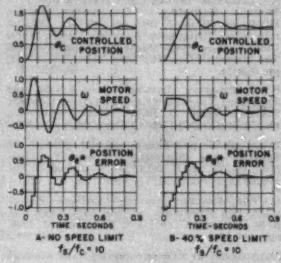


Fig. 16. Position control response with speed saturation

a step function. The position error of the sampled-data system does not deviate too much from the position error obtained by the continuous system. The continuous system can go from 0 to 100% position in one second with a peak position error of approximately 6% occurring 1/10 second after the application of the input signal. The sampled-data system error is approximately 7% occurring at the same time

In Fig. 15, the response of the system is shown for a constant acceleration input to the reference position, both for the continuous system and the sampled-data system. In the sampled-data system case, the sampling-frequency ratio has been reduced to 5 in order to show some deviation between the sampled-data system and the continuous system. With the sampling-frequency ratio at 10, there is no noticeable difference between the continuous system and the sampled-data system. However, in the case in which the sampling frequency has been reduced to 5 times the crossover frequency, there is some slight oscillation of the motor speed and the position error is slightly larger than that of the continuous system.

Nonlinear System Considerations

Up to this point in the analysis of the sampled-data system, the continuous portion of the system has been considered to be linear. However, in most practical problems, this is not true. In the case of the position servo being considered in this second example, the speed loop has a definite saturation level. If a step function is applied to the system, the speed will not rise to a high value and reduce the error signal as rapidly as shown in Fig. 13. The analog computer simulation of the sampled-data system can easily include the effects of this nonlinearity in the continuous portion of the system. The effect of speed saturation can be obtained by the insertion of limiting diodes across amplifier 4 of the simulation given in Fig. 12. The response of the sampled-data system to a step function input, both for the linear system and for the nonlinear system with motor speed saturation, is shown in Fig. 16. In Fig. 16(A), the system response is given for no motor speed saturation. In Fig. 16(B), a 40% speed limit is placed upon the speed servo. An interesting feature of the effect of the speed limit upon the behavior of the system can be seen by comparing Figs. 16(A) and (B). Since the insertion of the speed limit reduces the magnitude of the positive transient overshoot, the effective settling time for the system to respond to the step change in reference position is also reduced. In addition to the saturation type of nonlinearities in the system, there are other nonlinearities such as hysteresis and backlash in the position feedback which may be included.

Another area of sampled-data system investigation in which the analog computer technique method is useful is the consideration of the effect of the sampling duration time, In most methods of analysis, the sampling duration is considered to be instantaneous and that the hold circuit then maintains the sampled signal for the full portion of the sampling period. If the sampled-data system is considered as one in which the control information is obtained from a digital computer, the information is then sampled instantaneously and it is a true sampled-data system. However, one may have a system in which several channels of control are obtained by using a common analog device to obtain sum and difference, multiplication, or function generation functions. Such a multiplexed system may provide operation whereby the time-sharing of the common computer allows the sampling duration to be some portion of the sampling period. For example, if only two channels were involved, it would be possible for the system to be continuous for one half the time and to be holding during the remaining half sampling period that the other channel was using the common computer. One would expect such a quasi-sampled-data system to have performance characteristics somewhere between that of the continuous system and that of the pure sampled-data system. The results of applying a step function to a continuous system, a quasi-sampled-data system in which the sampling duration time to sampling period was one half, and the true sampled-data system are shown in Fig. 17. The positive-peak overshoot is found to be 36% for the continuous system, 45% for the quasi-sampled-data system, and 64% for the true sampled-data system The continuous portion of the system used in these three cases is that of the higher-order system considered as the second example. In both the quasi-sampled data system and the sampled-data system, a sampling-frequency to crossover-frequency ratio of 10 was used.

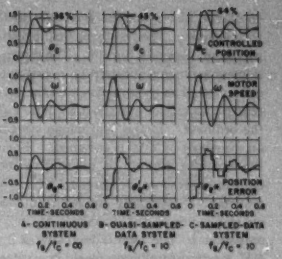


Fig. 17. Position control response for finite sampling duration

To illustrate the effect of the sampling duration upon the sampled-data system performance, a series of transient runs were taken in which this variation was noted. In this case, the first illustrative example of the second-order system with

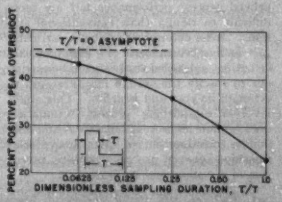


Fig. 18. Effect of sampling duration on second-order system response

45-degree phase margin was chosen as the example. The results of this investigation are shown in Fig. 18. The dimensionless sampling duration is logarithmically plotted as the abscissa, and the percent positive peak overshoot is plotted as the ordinate. In the continuous system an overshoot of 23% is obtained. When the sampling duration is reduced to one half, the overshoot increases to 30%. A reduction of sampling duration to one quarter increases the overshoot to 36%; reduction to one eighth increases it to 40%, asymptotically approaching a value of 46% when the sampling duration approaches zero.

The application of the analog computer method of studying sampled-data system of higher-order system including nonlinearities and the effect of the sampling duration are examples of other types of investigations which can be carried out using this technique.

Conclusions

From the results of the study of the simulation and analysis of sampled-data systems on the analog computer which has been briefly reviewed in this paper, several conclusions can be drawn concerning this method and its application.

There is good correlation between the results obtained from the analog computer simulation and the theoretical methods using the frequency domain and time domain approaches. In all cases in which the sampled-data system could be analyzed by the classic methods, a good check was obtained in both the transient-response and frequency-response areas of comparison. The application of the method to the very limited and somewhat idealized second-order system with the different values of linear system phase margin provides a convenient means of obtaining an order of magnitude correction for the effect of the sampling upon the continuous system performance.

Many control engineers are familiar with the application of analog computer techniques to the solution of both linear and nonlinear systems design and evaluation. The application of the analog computer to the simulation and analysis of sampled-data systems promotes an extension of the use of such systems, since technical feasibility can be demonstrated quite easily using the analog computer equipment. Because it is unnecessary to make limiting assumptions concerning such things as saturation, backlash, or other nonlinearities when performing a system analysis on the analog computer, the control engineer can obtain a much firmer indication of his system performance.

References

1. A Survey of Techniques for the Analysis of Sampled-Data Control Systems, **G. J. Murphy, R. D. Ormsby.** *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. AC-2, Feb. 1957, pp. 79–90 (contains extensive bibliography).

2. Automatic Feedback Control System Synthesis (book), **J. G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955, chap. 9.

3. The Analysis of Sampled-Data Systems, **J. R. Ragazzini, L. A. Zadeh.** *AIEE Transactions*, vol. 71, pt. II, 1952, pp. 225–32.

4. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, **W. K. Linvill.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1779–82.

5. Theory of Intermittent Regulation, **Y. Z. Tsypkin.** *Avtomatika i Telemekhanika*, Moscow, USSR, vol. 10, no. 3, 1949, pp. 189–224. English translation, Frequency Analysis (book), **Rufus Oldenburger, editor.** "Sampling Controls," Part 8, The Macmillan Company, New York, N. Y., 1956, pp. 309–24.

6. Correlation Between Root-Locus and Transient Response of Sampled-Data Control Systems, **Eliahu I. Jury.** *AIEE Transactions*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 427–35.

7. Additions to z-Transformation Theory for Sampled-Data Systems, **Gladwyn V. Lago.** *Ibid.*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 403–08.

8. The Analysis of Sampled-Data Servomechanism Performed on the IBM Type 650, **B. M. Tostanoski.** *Ibid.*, vol. 75, pt. I, Sept. 1956, pp. 446–50.

9. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, **G. Farmanfarma.** *Ibid.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 808–19.

System Considerations in Computer Control of a Semicontinuous Chemical Process

T. M. STOUT
ASSOCIATE MEMBER AIEE

Synopsis: Control systems incorporating a digital control computer can provide improved process performance but are more difficult to design than conventional systems. After determination that a particular process may warrant use of a digital control computer, a detailed study should be carried out to specify the system functions and to estimate its cost and benefits. This study involves establishing process boundaries, examination of process objectives and economics, development of quantitative relationships between process variables, invention of a system, and its evaluation.

To illustrate some of the steps in a study of this kind, the isomerization process for converting normal butane to isobutane is considered. The process is described, some basic relations are developed, and several possible functions of a digital control system are suggested. Particular attention is given to the problems connected with catalyst deactivation which arise from the semicontinuous nature of the process.

MANUFACTURING processes are designed and operated to produce a profit. This profit results from the creation, by a combination of physical and chemical transformations, of a product or products whose value exceeds the cost of the raw materials and their processing. Maximum profit, a major objective in business management, is obtained if correct decisions are made in the design and operation of the process.

Although governed by economic considerations, the control objectives for a given process are usually stated in more direct ways. A specified or maximum production rate must be realized; certain product specifications must be met; waste of fuel, water, and materials must be minimized; and conditions leading to equipment breakdowns, lost operating time, and unexpected maintenance costs, must be avoided. These objectives are, in turn, translated into specific operating conditions (flows, levels, temperatures, pressures, compositions, etc.) which should be maintained at various points in the process.

In general, therefore, the control problem can be stated in the following way:

Paper 58-1321, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted July 7, 1958; made available for printing September 4, 1958.

T. M. Stout is with The Thompson-Ramo-Wooldridge Products Company, Los Angeles, Calif.

The function of any control system, manual or automatic, is to make changes in certain controllable quantities to compensate for changes in other uncontrollable quantities so that process objectives are continuously met.

The varying and uncontrollable quantities may include feed rates and compositions, cooling water temperatures, fuel heating values, catalyst activity, and ambient conditions such as air temperature and wind velocity. Other quantities which influence process operation, although less rapidly, are variations in the market for the product, the price, the amount which can be sold, and the characteristics which are demanded by customers.

Philosophy of Computer Control Systems

If the uncontrollable quantities never changed, the operating procedures and conditions which result in maximum profit could be established and maintained for all time. Since frequent and inevitable variations do occur in quantities over which no control can be exercised, however, automatic control has been introduced. Single-variable control loops and ratio or cascade control systems involving several variables are commonly employed to counteract environmental disturbances. Although these controllers can do an effective job of maintaining a given set of process conditions, they do not necessarily provide optimum results.

In practice, the controllable-process variables are frequently maintained at fixed values even though it may be recognized that some of them should be changed to fit changes in uncontrolled quantities. In continuous catalytic processes, for example, it may be known that maximum yield is obtained by adjusting pressure and temperature to fit changes in feed rate, composition, and catalyst activity. Adjustments of this type are left to the process operator. To assist him in performing this assignment, various rules, charts, nomographs, and other aids are often supplied. However, these aids may be so simplified and condensed that process operation could only be partially optimized by their use. In addition, the operator is subject to fatigue and boredom, and he may not have time to make all the desired adjustments. For these reasons, it is sometimes preferable to have the operator make no adjustments or decisions. A. G. Green, discussing the question of whether color or alum should be added first in coloring paper stock, puts the matter very succinctly:[1]

> The practical way of running a mill is to have a fixed rule, one that is nearest right on the average, and one that will not involve a lot of men in the complexities of chemistry.

A digital control computer can be introduced which will accept readings from process instruments, perform complex calculations to determine the best settings for the controllable variables, and make the required adjustments automatically. Unlike the human operator, the control computer does not become tired or bored. It can handle many variables in a short time. The complexity of the control relationships is no longer limited by the fallibility of the operators but by the ingenuity of process engineers responsible for the operation of the system. For these reasons, the actions taken by a digital control computer can be right, not only on the average, but also in not-so-average situations. With essential compromises of present control procedures eliminated, more nearly optimum process operation is sure to result.

Process operation following installation of a digital control computer will not be perfect. The measuring and control devices associated with the computer are not perfect; the control functions built into the system at least initially, will not be ideal; and the computer itself is not infallible. Nevertheless, this new tool will permit significant improvements in process operation, improvements which can result in sufficient dollar benefits to justify the investment in control equipment.

System Engineering

These properties which make the digital control computer such a powerful instrument—its ability to exercise an integrated control of a process based upon interpretation, and manipulation of all important process variables—also make it more difficult to apply than conventional instrumentation. Before deciding to install a digital control computer, one must conduct a thorough analysis of the process. In this study, process objectives are defined and related to control actions which can be taken in the process. When this has been done, the general functions of a computer control system can be specified, its cost estimated, and its performance evaluated. These results then furnish the basis for a decision on the advisability of proceeding with detailed design, construction, installation, test, and start-up of such a computer control system.

Preliminary Questions

Since the feasibility study itself is a somewhat lengthy undertaking, it is advisable to establish in advance that a computer control system is likely to be useful for a particular process. This determination involves two questions:

> Can process operation be improved only by use of computer-control techniques?
>
> Will benefits from improved operation pay for the necessary equipment?

The first question is answered by examination of the general complexity of the process, the amounts and rates of variation of uncontrollable quantities, and the adequacy of present control techniques. The second question is answered by consideration of the volume and values of materials handled, the importance of product quality, and the operating costs.

In general, if several interrelated control decisions or adjustments must be made repeatedly in the face of significant variations in a number of uncontrollable variables and if enough dollars are involved (perhaps $3,000,000 in product value or operating costs per year), an affirmative answer to these questions can be expected. If examination of a particular process indicates that both answers are "yes," a closer investigation of the process is justified in order to specify more precisely what the control system should do, how to do it, what benefits will be provided, and what the system will cost.

Feasibility Study

Establishing the feasibility of a particular digital control system installation involves four main steps:

1. Fixing process boundaries.
2. Understanding the existing process.
3. Devising a new control system.
4. Evaluating the new control system.

Each of these steps includes a number of substeps which deserve specific description.

1. *Fixing process boundaries.* All of the important parts of the process must be included. At the same time, the area included must be small enough that an adequate study can be completed in a reasonable time.

The initial limits of the study may coincide with the geographical boundaries of the process. The boundaries may later be narrowed to specific parts of the process or expanded to include any adjacent processes which are relevant to the problem. In all cases, the boundaries must be chosen with care so that improve-

ments in one process do not inadvertently have adverse effects on other processes.

2. *Understanding the existing process.* The system engineer begins by collecting information on equipment layout and limitations, material flows and characteristics, and control objectives and procedures. He should be informed of any equipment modifications in the process which might affect his work. Data on flow rates and material characteristics must provide a clear picture of variability as well as normal values. Product specifications must be understood, and operating procedures should be examined for indications of interrelations between process variables.

The system engineer needs some data on material values and operating costs. It is particularly important that he have realistic information on those costs which vary with the way the process is operated. Fixed costs such as administrative charges, overhead, taxes, insurance, and advertising expense charged to the process generally will not concern him if they are truly independent of operating conditions.

An important and difficult aspect of any system study is the development of relations between the process variables. Operating data can usually be expressed by mathematical functions, developed with the aid of any applicable theory, which account for the most important relationships. These functions may be directly useful in control and will be used in evaluating the performance of any proposed control system.

In order to specify the frequency of measurement and control actions, and to insure system stability, it is necessary to determine how rapidly the independent variables change and how rapidly the process responds to changes in the independent and controlled variables. Process records and published literature may already be available which will provide adequate preliminary information, but plant tests may be necessary before a detailed system design can be fully completed.

3. *Devising a new control system.* The invention or design of a new computer-control system involves specification of the control relations that will be used, determination that necessary equipment is available, and statement of procedures to be followed under a variety of conditions.

To specify which variables will be measured and which will be controlled, consideration is given to process economics, product specifications, availability of suitable instruments, and the nature of equipment limitations and process bottlenecks. In the course of this consideration, the system engineer discovers whether the original statement of process objectives in economic terms should be directly involved in control decisions or adjustments, or whether it can be reduced to subsidiary objectives such as maximization of production rate or minimization of operating costs while observing one or more product-quality specifications. In order to use the digital-control system most effectively, it may be desirable to recommend changes in the layout of process equipment, new instruments or controllers, or additional storage, heating, or material-handling facilities.

Procedures for normal operation, start-up and shutdown, and emergency conditions should be considered. In particular, the respective roles of the digital control computer and the operator should be defined, and means of communication from one to the other should be described.

4. *Evaluation of the new control system.* A complete system evaluation involves an estimate of equipment and operating costs, a prediction of expected dollar benefits, and a combination of those figures in some kind of payout formula.

When the system is laid out, costs for new control equipment and related process modifications can be estimated fairly easily. Operating costs for the new system, including maintenance of the control system and process equipment, are somewhat more difficult to judge since little experience with such systems is available.

Dollar benefits can result from improved control of product quality or increased throughput or yield. If off-specification product must be discarded, rerun, or shipped back to the manufacturer at his expense, the dollar value of improved quality can readily be estimated. If better quality will gain some larger share of a limited market, it is difficult to predict the exact dollar advantage to be gained by improved control, even though real benefits will be realized. In some cases, a reduction in quality variations from day to day may permit a higher minimum quality and a reduction in average quality, with corresponding reductions in operating costs. Increased throughput or yield or both may result from utilization of process conditions which are closer to equipment limitations than is now feasible or from continuous adjustment of process operating points to optimum values.

Dollar savings can be realized by decreased use of materials, energy, or labor. Material and energy costs are decreased by more efficient use of chemicals (catalysts and solvents, for example), water, oil, gas, electricity, etc. Operating labor costs can, in principle, be reduced but this saving is usually insignificant in processes which are already highly automatic.

All of these areas should be examined. The expected benefits are predicted by comparing past performance of the process with the calculated performance that would have been obtained if the proposed system had been used. These calculations require determination on an hour-by-hour or day-by-day basis of the production rates, product quality, and operating costs for the same set of independent variables.

After the benefits and costs of the new system have been estimated, the attractiveness of investment in new control equipment can be evaluated in any of several standard ways: a simple ratio of investment to benefits, or more complicated formulas involving interest, depreciation, and tax rates.

Detailed Design and Installation

At the completion of the feasibility study, it will be established whether or not a computer control system can be devised which meets the following conditions:

1. Meets the economic and technical objectives of process operation.
2. Employs only available instruments for measurement and control functions.
3. Operates rapidly enough to provide accurate and stable control.
4. Satisfies restrictions imposed for the protection of equipment and personnel.

If such a system can be devised and if the estimated system costs and benefits are attractive, a decision may be made to go ahead with installation and operation of the system.

At this point, a significant amount of system engineering remains. The exact form of the control relations, the numerical values needed in the many equations, the frequency of measurement, the methods for dealing with erratic data, and other details must be settled. System dynamics must be carefully examined, and the computer program must be written and checked. All necessary procedures must be specified in detail; the extent and nature of system checks performed by the computer, and the actions to be taken in the event of a failure anywhere in the system, must be specified. Equipment must be selected, purchased or constructed, installed, and checked.

Maintenance procedures must be developed. Drawings, reports, and manuals must be prepared. In the course of this work, plans should be made for evolution of the system as new knowledge of the process and new instruments become available.

When the system is delivered and assembled, a period of testing and evaluation will begin. Inevitable difficulties will be uncovered and remedied by equipment or program changes. When everybody is satisfied with system performance, routine operation starts.

Types of Processes

Although no digital control systems have yet been installed for industrial processes, sufficient preliminary work has been done that some of the unique problems posed by batch, continuous, and semicontinuous processes can be discussed.

In batch processes, the flows of raw materials and products are deliberately intermittent. In a typical case, material is placed in a vessel and subjected to planned variations in temperature and pressure, perhaps accompanied by the programmed addition of other materials, which will lead to formation of the desired product. A combination of technical, economic, and practical considerations may lead to programs which give the desired product in a minimum time consistent with limitations on heat input, in a specified time with a minimum processing cost, or in some other optimum fashion. A digital control computer can be used to develop the optimum program for each batch, perhaps by accounting for variations in the initial conditions but perferably by also using data obtained as the batch is processed. The digital control computer, having developed the program, can also enforce the necessary pattern of variations on the controllable variables.[8]

In continuous processes, on the other hand, materials and products are expected to flow into and out of process vessels in steady streams. Fuel, steam, electricity, water, and chemical consumption rates are supposed to be relatively constant, as are levels, pressures, temperatures, compositions, and other conditions throughout the process.

All so-called continuous processes are, of course, shut down and started from time to time and, in addition, are subject to irregular upsets. During these periods, as well as periods of transition from one essentially steady operating condition to another, optimum control is based on the same considerations which govern optimization of batch processes. At other times, consideration of material, utility, and operating costs, product value, and the effect of operating variables upon product quality and yield permits calculation of the steady-state conditions which should be maintained to realize maximum profit, maximum production rate, or some other process objective. These conditions will, in general, shift with changes in the uncontrollable variables. The job of the computer-control system is to locate the optimum conditions and to enforce them by direct action on control valves or through adjustment of controller set points.

Semicontinuous processes also display both batch and continuous characteristics not sequentially but simultaneously. In a typical semicontinuous process, feed material is supplied continuously and catalyst is replaced or regenerated periodically. The production rate depends on catalyst activity which in turn decreases with time or the amount of material processed. Catalyst deactivation occurs at a rate that may vary with the severity of reaction conditions. Any attempt to increase the present production rate by, for example, increasing reaction temperature may accelerate catalyst deactivation and cause a net loss in total production over the time interval between catalyst replacements. Optimum process operation is therefore obtained by a judicious balance between product value and operating costs, neither maximum production rate nor minimum catalyst usage being the sole criterion.

In order to illustrate some of the steps in a feasibility study, to examine some of the special problems presented by semicontinuous processes, and to indicate ways in which a computer control system might be applied to such processes, the butane isomerization process has been taken as an example. Other processes could have been used with equal facility. The information on which the paper is based was obtained primarily from published sources.[8–11] Both data and space being unavailable, the paper gives only a partial view of the work that would be performed in a full-scale study of this process.

Butane Isomerization

Fundamentals

The two isomers (defined as substances having the same formula but different structures) of butane, C_4H_{10}, are shown in Fig. 1. The two compounds, sometimes written nC_4 and iC_4 for convenience, are gases at normal pressures and temperatures. Although they have the same number of carbon and hydrogen atoms, they have different physical and chemical properties. For example, at atmospheric pressure, normal butane boils at approximately 31 F (degrees Fahrenheit) and isobutane at 50 F. The branch-chain structure of isobutane makes it more active chemically.

Isomerization of normal to isobutane was introduced during World War II to increase the supply of isobutane available as feed stock to alkylation processes for making high-octane aviation gasoline. In a typical alkylation process, isobutane combines with butenes (C_4H_8) in the presence of sulfuric acid to produce isooctane (C_8H_{18}) which has an octane number of 100. Although the demand for aviation gasoline has dropped considerably since World War II, a gradually increasing octane requirement for motor gasoline has generated new interest in the process. Somewhat similar processes for direct production of motor gasoline stocks by isomerization of normal pentane (C_5H_{12}) and normal hexane (C_6H_{14}) are also assuming increased importance, five new or improved processes having become available within the last year or so.

Five processes have been used for butane isomerization, two of which are vapor-phase processes of the semicontinuous type. The other three are liquid-phase processes in which catalyst is continuously added and removed.[9]

Process Description

As with many other processes, the butane isomerization process consists of a

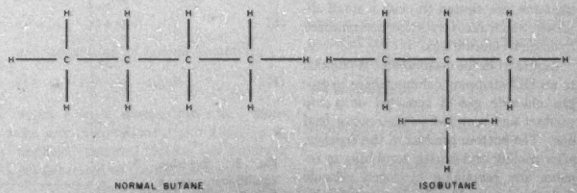


Fig. 1. Structural diagrams of the butane isomers

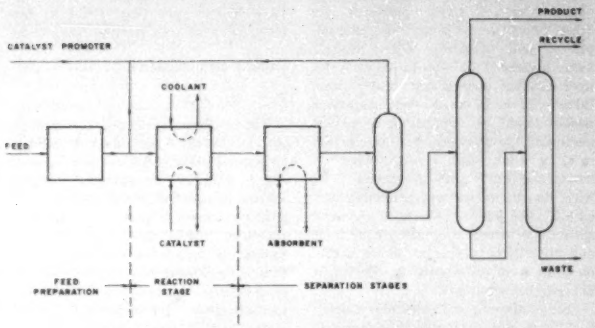


Fig. 2. Flow diagram for the butane isomerization process

feed-preparation stage, a reaction stage, and several separation stages, shown schematically in Fig. 2. In this diagram, a number of pumps, heat exchangers, and coolers have been omitted in the interests of simplicity.

The normal butane feed stream, partly fresh feed and partly recycle of unreacted material, is dried by washing with caustic soda and vaporized. The vapor and a small amount of dry hydrogen chloride gas, which acts as a catalyst promoter, are introduced into the reactors. The reactors may contain a single catalyst bed or may resemble heat exchangers with the catalyst placed in tubes surrounded by a circulating cooling medium. Several parallel reactors are generally used. The catalyst is aluminum chloride supported on alumina; fresh catalyst contains 15–25% aluminum chloride. The reaction is exothermic, about 76 Btu being evolved per pound of isobutane formed.

The material leaving the reactor flows through absorbers which remove small amounts of volatilized aluminum chloride which would otherwise foul subsequent equipment. After water cooling and refrigeration, the material enters an accumulator which is maintained at about 50 F. Uncondensable gases formed in the reactors, chiefly hydrogen and methane, are vented through a small absorber which recovers some uncondensed hydrogen chloride gas.

Liquid from the accumulator is pumped to an HCl stripping column where hydrogen chloride gas is removed as a top product and recycled to the reactor feed line. The bottom product of the stripper, after cooling and caustic scrubbing to remove any remaining hydrogen chloride and aluminum chloride, is separated by fractionation.

The overhead product from the column is the isobutane product which will contain small amounts of ethane and propane formed in the reaction. The bottom product is largely unconverted normal butane, with some isobutane and pentanes also present. The heavier hydrocarbons are removed in another stage.

Operating Problems

Increased temperature has a marked effect on the reaction rate; according to one reference the reaction rate doubles for a 15 F rise in temperature.[6] On the other hand, the equilibrium concentration of isobutane in an nC_4–iC_4 mixture decreases with temperature. Although these conflicting trends indicate the existence of an optimum temperature, other considerations help to determine the temperature employed.

Catalyst activity decreases gradually due to loss of aluminum chloride by sublimation and coating products of side reactions, both effects being more serious at higher temperatures. The sublimation rate increases with feed rate, but decreases with reaction pressure. It has been reported[4] that the amount of low-boiling side products is more than doubled by increasing the temperature from 210 F to 250 F. The concentration is usually about 2%.

Since higher pressures permit increased reaction times for a given feed rate or an increased feed rate for the same reaction time, with no significant effect on reaction equilibrium, it is desirable to use the maximum practical pressure at all times. The pressure limit is fixed by equipment restrictions and by the dew point of the reacting material, since condensation of liquids on the catalyst should be avoided.

The catalyst activity is also reduced by impurities in the feed. Sulfur compounds would be harmful but are usually removed in preceding refining processes. Any remaining traces are caught in the initial caustic wash which is intended primarily to remove water that would otherwise damage the catalyst and also create corrosion problems. Olefins and pentanes in the feed would likewise have adverse effects on the catalyst and are therefore removed in preceding processes.

The conversion of normal butane to isobutane increases as the hydrogen chloride content of the feed mixture increases, although not in a linear manner. Starting at 1–2% by weight with fresh catalyst, the HCl concentration may be increased to 16–20% before the catalyst is discarded. Gradual increases in temperature are used as another means of maintaining conversion.

Other conditions being fixed, the conversion decreases as the feed rate increases since the effective reaction time is reduced.

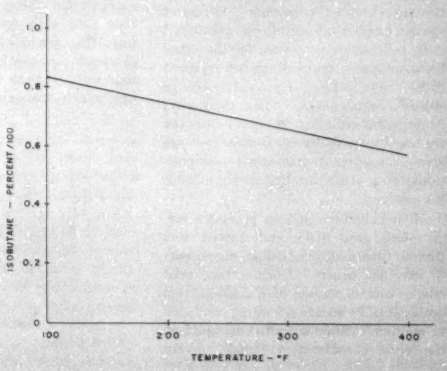


Fig. 3. Reaction equilibrium as a function of temperature

The production rate is directly proportional to feed rate, however, and a rate for maximum production rate might be anticipated.

Fixing Process Boundaries

Assuming that the sources of normal butane feed will not be affected by control actions in the isomerization process, a boundary can be established on the input side of the process. Data on variations in feed rate and compositions, not available in the published literature, are necessary for a complete study.

If isobutane of the required purity is always produced, it appears that the effects of subsequent processes will be felt by the isomerization process primarily as changes in demand for its isobutane product. Depending on the circumstances, the isomerization process may be asked to deliver at some definite rate or to deliver all the isobutane that can be produced. Again, necessary data are not available to characterize conditions at this process boundary.

Normally, the feed preparation and product fractionation stages would be unlikely to require any adjustment by a computer-control system. In this case, the process boundaries can be narrowed and attention concentrated on the reaction stage.

Understanding the Process

The primary measure of process performance is conversion, from which production rate and yield can be calculated. Because of the simplicity of the process, a theoretical relationship between conversion and the process variables can be developed which will serve as an adequate approximation to the facts of actual operation.

To simplify things, assume that the isobutane content of the feed is 100 x_o% and the normal butane content is 100 $(1-x_o)$%. Given sufficient time, the reaction product would reach the equilibrium composition shown in Fig. 3 with 100x_e% isobutane. The corresponding conversion is

$$c_e = \frac{x_e - x_o}{1 - x_o} \tag{1}$$

In practice, a smaller isobutane concentration is obtained, denoted as 100x %. The actual conversion under practical conditions is then

$$c = \frac{x - x_o}{1 - x_o} \tag{2}$$

Leaving the reactor, the amount of isobutane is

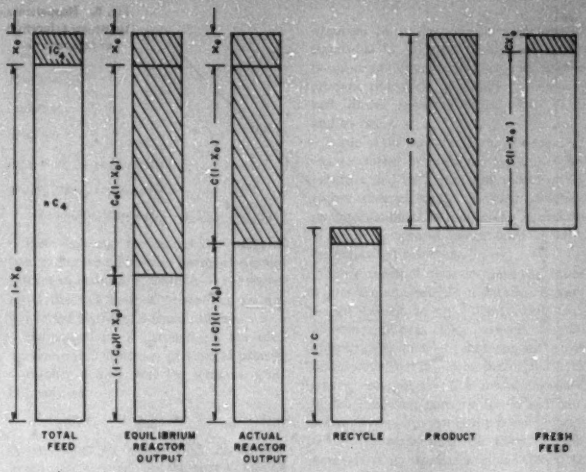


Fig. 4. Assumed compositions of various process streams

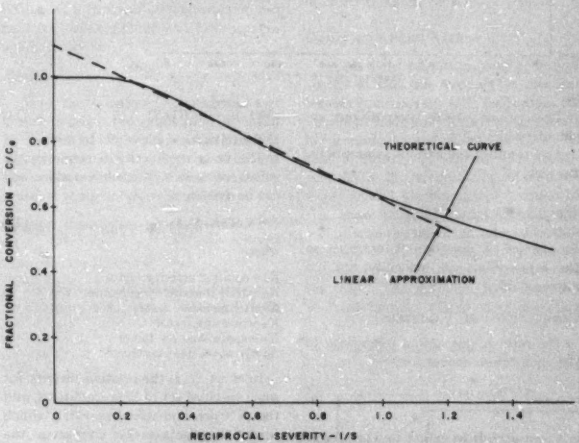


Fig. 5. Fractional conversion as a function of reciprocal reaction severity

$$xF = [c(1-x_o) + x_o]F \tag{3}$$

and the amount of normal butane is

$$(1-x)F = [(1-c)(1-x_o)]F \tag{4}$$

where F is the reactor feed rate. Since the sum of equations 3 and 4 is F, it is implicitly assumed that no side products are generated in the reaction.

If it is further assumed that the fractionation system delivers a pure isobutane product and leaves as recycle a material having the same composition as the reactor feed;

$$P = \text{product rate} = cF \tag{5}$$

$$R = \text{recycle rate} = (1-c)F \tag{6}$$

$$F' = \text{fresh feed rate} = F - R = P = cF \tag{7}$$

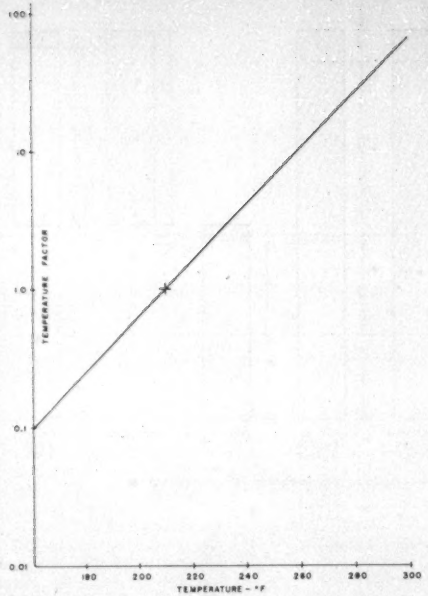


Fig. 6. Temperature factor as a function of temperature

The various stream compositions are summarized in Fig. 4.

As a result of the various assumptions, the yield is

$$Y=\frac{P}{F'}=1.00 \qquad (8)$$

In practice, Y may be 0.95 or more, so the assumptions are evidently not too unreasonable.

Correlation of Variables

The ratio of the actual conversion to the equilibrium conversion,

$$c'=\frac{c}{c_e}=\frac{x-x_0}{x_e-x_0} \qquad (9)$$

is a measure of the extent to which equilibrium is attained and can be written (see Appendix I)

$$\ln\frac{1}{1-c'}=S \qquad (10)$$

where

c' = fractional conversion, per cent/100
S = reaction severity

A plot of c' as a function of $1/S$ is given in Fig. 5. It will be noted that c' increases as $1/S$ decreases, i.e., as S becomes larger. For some purposes, a linear approximation

$$c'\cong 1.10-0.47(1/S) \qquad (11)$$

to the theoretical curve will be useful.

The reaction severity S combines the effects of several process variables and can be written

$$S=K_cK_QK_\theta K_\pi K_F S_B \qquad (12)$$

where

K_c = catalyst activity factor
K_Q = HCl concentration factor
K_θ = temperature factor
K_π = pressure factor
K_F = space velocity factor
S_B = base reaction severity

In effect, S_B is the reaction severity for some specified set of base conditions, and the K's are dimensionless ratios which permit the effects on S of changes in the other variables to be easily calculated. If S_B is assigned arbitrarily, K_c can be determined from a knowledge of c' and the other factors; conversely, K_c can be chosen arbitrarily for some set of conditions, S_B evaluated from c' and the other factors, and then this value of S_B used in subsequent calculations to study variations in K_c. The latter approach will be adopted, S_B being so chosen that K_c is approximately unity for fresh catalyst.

The catalyst activity factor K_c is proportional to catalyst activity.

The HCl concentration factor K_Q increases with the amount of hydrogen chloride gas in the reactor feed. The exact relation is nonlinear and cannot be evaluated from published data. For convenience, K_Q may be taken equal to unity for an HCl concentration of 10%.

The temperature factor K_θ combines the effects of temperature (denoted by θ so that T can be used as a symbol for time) on reaction rate constant, equilibrium isobutane concentration, and space velocity. As shown in Fig. 6, the temperature factor has arbitrarily been made unity at 210 F and, as stated earlier, doubles in value for increases of 15 F.

The pressure factor can be taken arbitrarily as unity as 250 psig (pounds per square inch gage), and expressed as

$$K_\pi=\frac{\pi+14.7}{264.7}$$

where π is the reaction pressure (psig).

The space velocity is ordinarily expressed as the liquid volume of feed per hour per volume of catalyst and varies in typical isomerization processes from 0.5 to 1.0. The space velocity factor K_F is the reciprocal of the space velocity.

As an illustration of the use of these equations, suppose that:

$x_0=0.02$
$\theta=210$ F
$K_c=1.00$
$K_Q=1.00$
$K_\theta=1.00$
$K_\pi=1.00$
$K_F=1.00$
$S_B=1.7$

From equation 12, we find that $S=1.7$ which gives, using equation 10, $c'=0.817$.

From Fig. 3, the equilibrium concentration of isobutane at this temperature is

$x_e=0.738$ so that

$$c_e=\frac{0.738-0.02}{1-0.02}$$
$$=0.733$$

and

$$c=(0.733)(0.817)$$
$$=0.600$$

or 60%.

If the reaction severity should be cut in half by doubling the feed rate or by a decrease in catalyst activity, with no change in the other conditions, a similar series of calculations gives

$S=0.85$
$c'=0.573$
$c=0.42$

Catalyst Deactivation

The loss of catalyst activity is due to poisoning, coating by products of side re-

actions, and sublimation of aluminum chloride. These effects probably depend directly on the total amount of feed processed since the catalyst was new, weighted by a function of temperature, pressure, HCl concentration, and perhaps other variables. If so, the catalyst activity factor can be expressed as

$$K_c(t) = K_{co} - \int_0^T \psi(G, \theta, \pi) F dt \quad (13)$$

where $K_{co} = K_c(0)$ and $\psi(G, \theta, \pi)$ is a function to be determined from operating data.

Process Dynamics

No specific data are available on the magnitudes or rates of changes of the independent variables or the speed of process response to changes in independent or controlled variables. It is clear, nevertheless, that an important effect is the time required for material to travel through the process. This delay or lag, perhaps 10 to 20 minutes, should be considered in calculating conversion figures. If tight control of production rate or conversion is found desirable, the compositions of the reactor feed and output should be measured on the reactor side of the feed preparation and product separation stages, respectively, to reduce this delay.

Process Economics

An approximate dollar balance for the process, based on Fig. 7, leads to the equation

$$D = \int_0^T [Pv_P + Rv_R + Wv_W - Fv_F - Gv_G - Hv_H] dt - n_C D_C - n_A D_A - D_O \quad (14)$$

where

D = return from process operation, dollars
D_C = cost of catalyst change (labor, material, and lost production during change), dollars
D_A = cost of absorbent change, dollars
D_O = fixed costs, dollars per day
P = production rate, barrels per day
R = recycle rate, barrels per day
W = waste rate, barrels per day
F = reactor feed rate, barrels per day
G = HCl rate, pounds per day
H = coolant rate, gallons per day
v_P = product value less proportional costs for cooling, separation, etc., dollars per barrel
v_R = recycle value less proportional costs for cooling, separation, etc., dollars per barrel
v_W = waste value less proportional costs for cooling, separation, etc., dollars per barrel
v_F = feed value plus proportional costs for drying, heating, etc., dollars per barrel
v_G = cost of HCl gas, dollars per pound
v_H = cost of coolant, dollars per gallon
t = time, days
T = total operating period, days
n_C = number of catalyst changes
n_A = number of absorbent changes

If it is assumed, as before, that

$$P = cF \quad (5)$$

$$R = (1-c)F \quad (6)$$

and that

$$Wv_W = 0$$

equation 14 reduces to

$$D = \int_0^T [cF(v_P - v_R) - F(v_F - v_R) - Gv_G - Hv_H] dt - n_C D_C - n_A D_A - D_O \quad (15)$$

Here $(v_P - v_R)$ represents the value of converting an additional barrel of normal butane to isobutane, while $(v_F - v_R)$ represents the net cost of processing an unconverted barrel of normal butane.

By equations 9 through 12, the conversion c is a function of several process variables which can be written symbolically as

$$c = \phi(K_c, G, \theta, \pi, F, z_0) \quad (16)$$

As indicated in equation 13, K_c at any given time depends on the past values of G, θ, π, and F.

The amount or circulation rate of the coolant is a function of temperature and feed rate, among other variables, written symbolically as

$$H = \xi(\theta, F) \quad (17)$$

Since the reaction is exothermic and therefore self-heating, less coolant should be needed at high temperatures and H will decrease with θ. The cost of cooling the reactors will therefore decrease as the reaction temperature is increased.

Devising a Control System

Control Objectives

The objective of process operation will be to obtain a maximum value of D from equation 14, subject to the conditions of equations 13, 16, and 17 and any additional conditions such as maximum usable product rate, feed availability, temperature limits, pressure limits, HCl rate, and so on.

Some of the constraints may be real and exact. Other limitations, such as temperature and HCl rate, may reflect qualitative understandings of these variables on catalyst deactivation or the formation of side products and may, at least initially, be somewhat arbitrary.

Since essential data on these and other points are unavailable, a detailed system design cannot be developed. One can, however, discuss some aspects of the optimization problem and suggest some functions which might be performed by a control computer. The items discussed in the following paragraphs do not constitute a complete list of the possible functions. It might also be noted in advance that no single function is likely to justify the total investment in control equipment, any payout being the result of several kinds of dollar benefits.

Control of Feed Rate

Calculations using equations 5, 9, and 10, show that the production rate increases continuously with feed rate. At very high feed rates, the conversion tends to decrease approximately as $(1/F)$ so that a limiting production rate is approached. If the unconverted material is all recycled, examination of equation 15 will show that an optimum feed rate will be found at which the value of further production will be balanced by the increased cost of processing unconverted material. The optimum feed rate is a complex function of the other operating

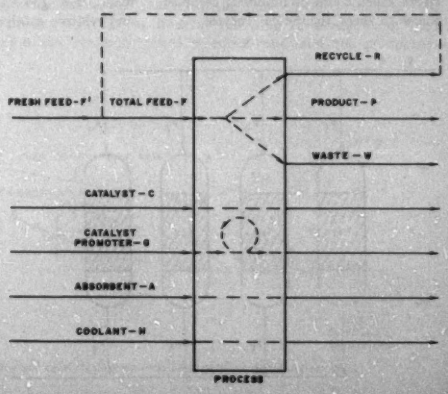


Fig. 7. Schematic diagram of process for use in economic analysis

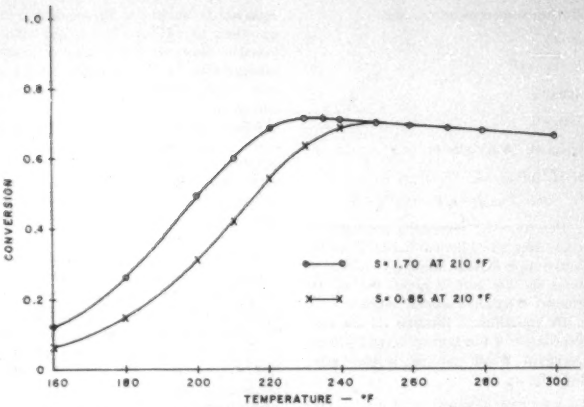


Fig. 8. Conversion as a function of temperature at two values of reaction severity

variables and operating costs which, under normal conditions, will probably exceed the available feed supply. At some times, however, it could be advantageous to discard part of the available feed in order to increase the dollar return from process operation.

Control of Temperature

Calculations using equations 9, 10, and 12 and Figs. 3 and 6 show a temperature at which maximum conversion is obtained. This point arises from an increasing reaction rate and a decreasing reaction equilibrium as the reaction temperature is increased. Typical results of these calculations are shown in Fig. 8; these curves have the same general shape as a conversion curve given in reference 8 for a continuous mixed-phase process, the only curve which could be found for comparison with these theoretical curves.

The temperature for maximum conversion is a function of the other variables which affect reaction severity and shifts with changes in catalyst activity, feed rate, reaction pressure, and HCl rate. Although the reaction temperature can probably be picked with only slight regard for cooling costs, the effects on catalyst activity and side-product formation must be considered. The control system may therefore maintain the temperature for maximum conversion only part of the time and under special circumstances.

Control of Feed Division

If parallel reactors are used which contain catalyst at different relative catalyst activities, as shown in Fig. 9, some increase in total production can be achieved by a judicious division of the available feed. If F_i and K_{ei} denote the feed rate and relative catalyst activity in the ith reactor and if other conditions are the same in the reactors, a simple calculation (see Appendix II) shows that the total production is a maximum if

$$\frac{F_1}{K_{e1}} = \frac{F_2}{K_{e2}} = \frac{F_3}{K_{e3}} = \cdots = \frac{F_i}{K_{ei}} \qquad (18)$$

Employment of this control function would require a knowledge of the individual catalyst activities. As previously noted, these quantities can be deduced from the conversion and other variables affecting S in equation 12. The individual reactor conversions could be obtained directly from composition measurements or, if the necessary analysis instruments proved too expensive for installation on every reactor outlet, indirectly from heat balances involving only flow and temperature measurements.

Scheduling of Catalyst Changes

The amount of catalyst used has significant effects on gross production and operating costs. A large number of catalyst changes (n_C) will increase the average catalyst activity and therefore the average production rate but will also increase the total catalyst cost. An optimum value of n_C can therefore be expected for any specified period.

A further complication arises from the necessity for shutdowns which are scheduled at regular intervals for inspection and preventive maintenance of process equipment. Since catalyst changes are conveniently or perhaps necessarily made during the shutdowns, the catalyst change schedule is not completely arbitrary. In order to avoid a catalyst change just prior to a shutdown, it is common practice to divide the period between shutdowns into an integer number of runs of roughly equal duration, say two, three or four per year.

If the catalyst activity does not decline in identical fashion from one load to the next, the run lengths should be ad-

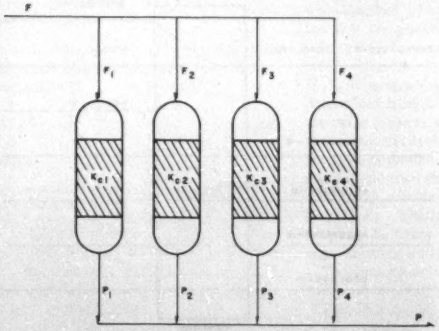
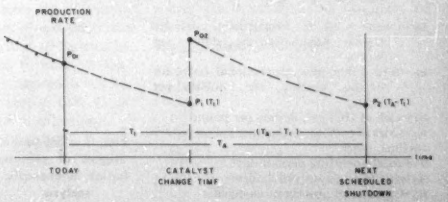


Fig. 9 (left). Parallel arrangement of reactors

Fig. 10 (below). Prediction of catalyst change time



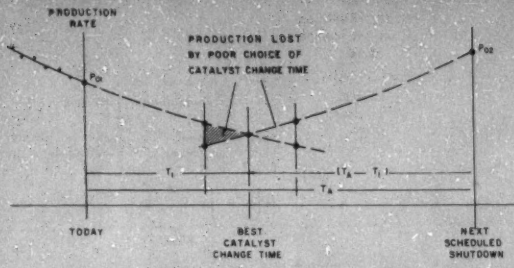


Fig. 11. Alternate construction for prediction of catalyst change time

justed accordingly. Runs with slow decline rates should be longer than those with high decline rates.

Selection of the optimum number and times of catalyst changes can be made in the following way:

1. Assume a number of runs to be made before the next scheduled shutdown.
2. By extrapolating the current deactivation rate for the run in progress and assuming normal deactivation rates for future runs, predict the optimum change times.
3. Knowing the run lengths, compute the expected dollar return over the time available.
4. Repeat for other numbers of runs and pick the number giving the largest dollar return.

Data for these calculations include curves of catalyst activity as a function of time or amount of product. In general, this information will not be available when a computer-control system is initially installed, and one of its first jobs will therefore be collection of the necessary data for prediction of catalyst change times. The criteria, however, can be specified in advance.

Final Run

Just before a scheduled shutdown, no catalyst changes need be considered. The objective at this time may be maximum production or dollar return. Consideration of equations 10, 12, 13, 14, 16 and 17, in combination with other constraints, should furnish a program of operation which will achieve the specified objective in the face of the various interacting effects.

One Catalyst Change

If enough time remains before the scheduled shutdown, one catalyst change may be advisable. For ease of discussion, it is assumed that the operating conditions are fixed. This assumption eliminates the variable operating costs and reduces the problem to selection of a catalyst change time which gives the maximum total production in the available time. If T_A is the available time and T_1 is the unknown date for the catalyst change, the total production is

$$Q(T_A) = Q_1(T_1) + Q_2(T_A - T_1) \qquad (19)$$

where

$Q_1(t)$ = total production after t days for the first run, and
$Q_2(t)$ = total production after t days for the second run

Differentiating with respect to T_1 and noting that dQ/dt is the production rate P, we find that the best catalyst change time is determined from the relation

$$P_1(T_1) = P_2(T_A - T_1) \qquad (20)$$

where

$P_1(t)$ = production rate after t days for the first run, and
$P_2(t)$ = production rate after t days for the second run

This criterion can be interpreted graphically with the aid of Figs. 10 and 11. In Fig. 10, the production rates for the two runs are plotted as functions of time; the first run starts at a production rate P_{01} and decreases to $P_1(T_1)$, while the second run starts at P_{02} and decreases to $P_2(T_A - T_1)$. When T_1 is chosen so that equation 20 is satisfied, the total production over the whole period is a maximum; this fact is demonstrated in Fig. 11. Here the production rate for the second run is plotted backward, and the intersection of the two curves defines the best catalyst change time. Changing catalyst either sooner or later would result in loss of potential production indicated by the shaded areas.

This criterion compares, in effect, the present trend in production rate or catalyst deactivation with a typical or normal curve. If the current catalyst is declining less rapidly than usual, the catalyst change will be delayed. Conversely, if the current catalyst is declining more rapidly, the catalyst change will be made sooner. These results, which fit common-sense notions of effective catalyst usage, are shown in Fig. 12.

Two or More Catalyst Changes

When two or more catalyst changes are assumed to occur in the available operating time T_A, as shown in Fig. 13, the criterion for selection of the best catalyst change times is essentially unchanged. With no knowledge of catalyst deactivation in future runs, a safe course is to assume that future runs will be identical and typical. The total production will then be

$$Q(T_A) = Q_1(T_1) + nQ_2\left(\frac{T_A - T_1}{n}\right) \qquad (21)$$

where n is the number of runs following the next catalyst change. Differentiating with respect to T_1 and equating to zero as before, it is now found that the best change times are given by

$$P_1(T_1) = P_2\left(\frac{T_A - T_1}{n}\right) \qquad (22)$$

That is, based on the information available at the time of the prediction, the interval following T_1 should be divided into n equal runs, and the production

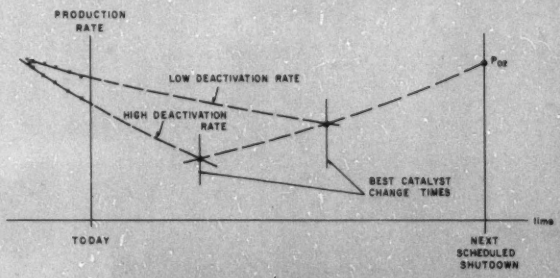


Fig. 12. Shift of catalyst change time with catalyst deactivation rate

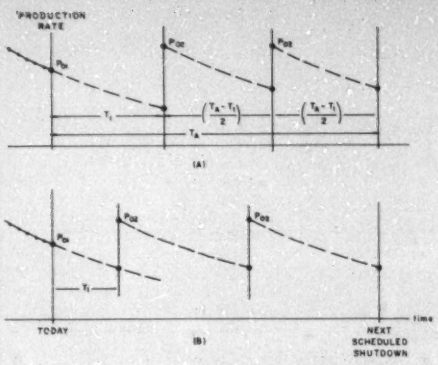


Fig. 13. Prediction of catalyst change times

A—Nonoptimum schedule
B—Optimum schedule

rates at the ends of all runs should equal the rate at time T_1, as shown in Fig. 13(B). After the catalyst change is made at T_1, and some data are collected during the next run, the same criterion may be applied again to predict the next catalyst change time. This time will not in general be $1/n$th of the remaining time, but will shift forward or backward as suggested by Fig. 12 to fit actual operating data.

Other Functions

In addition to other control functions, such as adjustments of the HCl rate, reaction pressure, or other variables in the feed preparation or separation stages, a computer-control system would undoubtedly have a number of scanning, alarm, and logging functions.

Variables not essential to the control program but nevertheless important could be received by the computer and processed in any desired manner. All of the input variables and various computed quantities, including the control outputs, could be checked against prescribed limits. In addition, trends and combinations of conditions denoting imminent or actual hazards to equipment or personnel could be detected. Finally, variables and computed quantities (including integrated and average flow rates) could be printed out for use by the process engineering and accounting departments.

The final determination of functions which should be performed and operating procedures to be followed under normal and emergency conditions is part of the detailed system design. In the feasibility study, these aspects must receive sufficient attention that an effective system can be assured, costs estimated accurately, and benefits assessed.

Conclusions

Design of a control system incorporating a digital control computer is a time-consuming but stimulating and not impossible task. One of the most difficult steps—development of quantitative relationships between process variables and objectives—has been partially negotiated in this paper for a relatively simple process. More complicated processes, for which more complete data are available, are being tackled in the same manner with equal success. Solutions to other problems encountered in system design are likewise being developed. It appears to be only a matter of time before such systems will be installed and operating experience reported.

Appendix I. Derivation of the Conversion Equation

Let x denote the concentration of isobutane at any time in an nC_4–iC_4 mixture. The rate of formation of isobutane can be expressed as

$$\frac{dx}{dt}=k_1(1-x)-k_2x \tag{23}$$

$$=k_1-(k_1+k_2)x \tag{24}$$

If it is defined that

$x(O)=x_o$ = initial isobutane concentration

and

$x(\infty)=x_e$ = equilibrium isobutane concentration

the solution of equation 24 can be written

$$x=(x_o-x_e)e^{-(k_1+k_2)t}+x_e \tag{25}$$

where

$$x_e=\frac{k_1}{k_1+k_2} \tag{26}$$

By introducing the quantities

$$c=\frac{x-x_o}{1-x_o}=\text{conversion} \tag{27}$$

and

$$c_e=\frac{x_e-x_o}{1-x_o}=\text{equilibrium conversion} \tag{28}$$

and performing some algebraic manipulations, equation 25 can be put into the form

$$ln\left[\frac{1}{1-c'}\right]=(k_1+k_2)t \tag{29}$$

$$=\frac{k_1t}{x_e} \tag{30}$$

$$=S \tag{31}$$

where

$$c'=\frac{x-x_o}{x_e-x_o}=\text{fractional conversion} \tag{32}$$

The forward reaction rate k_1 and the equilibrium concentration x_e are functions of temperature. The effective reaction time t can be expressed as

$$t=\frac{K\pi C}{\theta F} \tag{33}$$

where

K = a constant
π = pressure, psia
C = catalyst weight, pounds
θ = temperature, degrees Kelvin
F = feed rate, barrels per day

Substituting the various quantities in equation 30, combining the functions of temperature into a single factor, and multiplying and dividing by suitable constants, it is possible to express the reaction severity as

$$S=K_cK_\theta K_\pi K_p S_B \tag{34}$$

The factors in equation 34 are discussed in the text.

Appendix II. Feed Division in Parallel Reactors

To simplify this derivation, it is assumed that all conditions except feed rate and catalyst activity are the same in all reactors. Equation 11 can then be rewritten

$$c_i'=1.10-m\frac{F_i}{K_{ci}} \tag{35}$$

where

m = a constant
F_i = feed rate for the ith reactor
K_{ci} = catalyst activity factor for the ith reactor

The actual conversion in the ith reactor is

$$c_i=c_ec_i'$$

$$=c_e\left[1.10-m\frac{F_i}{K_{ci}}\right] \tag{36}$$

where the equilibrium conversion c_e (which depends on x_o and θ) is common to all reactors.

The total production rate for j reactors is

$$P=\sum_{1}^{j} c_i F_i \tag{37}$$

$$=\sum_{1}^{j} c_e\left[1.10-m\frac{F_i}{K_{ei}}\right]F_i \tag{38}$$

which is to be maximized subject to the condition that

$$F=\sum_{1}^{j} F_i \tag{39}$$

or

$$F-\sum_{1}^{j} F_i=0 \tag{40}$$

By the method of Lagrangian multipliers,[12] the function

$$P+\lambda\left[F-\sum_{1}^{j} F_i\right] \tag{41}$$

is differentiated with respect to the various F_i from F_1 to F_j. The resulting quantities are set equal to zero, giving j equations of the form

$$c_e\left[1.10-2m\frac{F_i}{K_{ei}}\right]-\lambda=0 \tag{42}$$

Eliminating λ, it is found that

$$\frac{F_1}{K_{e1}}=\frac{F_2}{K_{e2}}=\ldots=\frac{F_j}{K_{ej}} \tag{43}$$

Further algebraic manipulations yield

$$\frac{F_i}{K_{ei}}=\frac{\sum_{1}^{j} F_i}{\sum_{1}^{j} K_{ei}} \tag{44}$$

or

$$F_i=\frac{K_{ei}}{\sum_{1}^{j} K_{ei}}F \tag{45}$$

which permit calculation of the optimum flow to each reactor if the total flow and the relative catalyst activities for all reactors are know. Alternately, substitution of equation 43 into equation 36 shows that the conversions should be identical in all reactors at optimum conditions.

References

1. Beating and Refining, A. B. Green. *Pulp and Paper Manufacture*, New York, N. Y., vol. 2, 1951, p. 221.

2. A Report on Optimizing Control of a Chemical Process, D. P. Eckman, I. Lefkowitz. *Control Engineering*, New York, N. Y., vol. 4, Sept. 1957, pp. 197–204.

3. Petroleum Refinery Engineering (book), W. L. Nelson. McGraw-Hill Book Company, Inc., New York, N. Y., third edition 1949, pp. 662–63.

4. The Chemical Technology of Petroleum (book), W. A. Gruse, D. R. Stevens. *Ibid.*, second edition 1942, pp. 56–63, 446–47.

5. Butane Vapor Phase Isomerization. *Petroleum Refiner*, Houston, Tex., vol. 33, Sept. 1954, p. 204.

6. The Shell Vapor-Phase Isomerization Process for the Production of Isobutane, H. A. Chaney, C. L. Raymond. *Transactions*, American Institute of Chemical Engineers, New York, N. Y., vol. 42, 1946, pp. 595–609.

7. Isomerization of Light Hydrocarbons, S. F. Perry. *Ibid.*, pp. 639–46.

8. Butane Isomerization—U.O.P. Process, J. A. Chenicek, J. O. Iverson, R. E. Sutherland, P. C. Weinert. *Chemical Engineering Progress*, New York, N. Y., vol. 43, May 1947, pp. 210–18.

9. Isomerization, R. C. Gunness. *Progress in Petroleum Technology*, American Chemical Society, Washington, D. C., Aug. 7, 1951, pp. 109–19.

10. Isomerization, S. F. Perry. *Industrial and Engineering Chemistry*, Washington, D. C., vol. 49, Sept. 1957, pp. 1532–33.

11. Normal Butane Catalytic Isomerization. *Oil and Gas Journal*, Tulsa, Okla., vol. 54, Mar. 19, 1956, p. 182.

12. Differential and Integral Calculus—Vol. II (book), R. Courant. Interscience Publishers, New York, N. Y., 1949, pp. 188–203.

Optimum Response of Discontinuous Feedback Control Systems

F. W. NESLINE, JR.
ASSOCIATE MEMBER AIEE

THE CONCEPT of "optimum" lies at the heart of all sophisticated design work. The intelligent designer always wants to know which design meets a given set of conflicting conditions and constraints in the best possible, or optimum, way. This is especially true in feedback control system engineering, where, in order to describe optimum quantitatively, its meaning is narrowed to include only system performance characteristics. Usually a performance index is chosen, and the optimum system is defined as the system that minimizes (or maximizes) the performance index for the given set of constraints.

In the field of relay, or bivalued discontinuous feedback control systems, this approach has been applied by McDonald[1] and extended by Bushaw[2] and others.[3] The performance index is chosen to be the time required by the system to reduce the error, and all of its derivatives, to zero after the occurrence of an input to the system. The system is then chosen, within certain imposed constraints, to minimize the performance index for a given system input or class of inputs. The purpose of this paper is as follows: 1. to discuss the effect of the input on optimization of discontinuous systems, and 2. to discuss optimization when the discontinuous element is a general multiposition device.

The Discontinuous-Feedback Control-System Problem

The general system under investigation is shown in Fig. 1. The general problem is: What characteristic must the computer exhibit in order to minimize the performance index when the system dynamics, the discontinuous element, and the input are given? In order to accomplish this task, it is necessary to visualize the totality of dynamical solutions for the error. The phase-space technique proves most useful in this respect.

This problem has been solved for a general set of system dynamics, for a discontinuous device that is an instantaneous 2-position switch (Fig. 2), and for a certain class of inputs.[1–3] To illustrate this solution, consider the system to be linear and underdamped second order. After nondimensionalizing (see Appendix), the error differential equation is

$$e''+2De'+e=r''+2Dr'+r-\operatorname{sgn} f \tag{1}$$

The force free solutions to equation 1 are obtained by setting all driving terms to zero. These solutions depend upon initial conditions only and are a family of nonintersecting spirals about the origin of the

Paper 58-1192, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted March 13, 1958; made available for printing August 5, 1958.

F. W. Nesline, Jr., is with the Massachusetts Institute of Technology, Cambridge, Mass.

This paper is part of a dissertation presented to Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research was aided by a grant from the Sigma-Xi-RESA Research Fund. Thanks are due Prof. Felix Zweig of Yale University for initiating this investigation and for supplying an outstanding example of the spirit of university research by his personal curiosity and enthusiasm for studies of unusual phenomena.

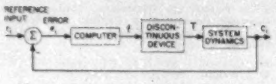


Fig. 1. Block diagram of control system

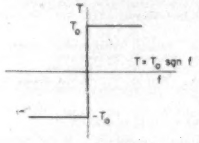


Fig. 2. Basic discontinuous device

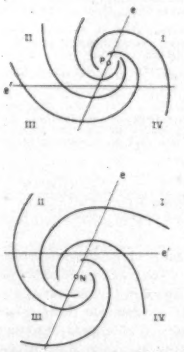


Fig. 3. Phase portraits for P and N singularities

e—e' plane.[4] If the reference input is kept zero, the sgn f term changes this phase portrait to two sets of spirals, one about the point $e=1$, called the P family, and the other about the point $e=-1$, called the N family, as shown in Fig. 3. In this figure and those that follow, the e and e' axes are not at right angles to each other in other to make the spirals logarithmic.[4,5] Since the phase trajectories do not intersect, only one trajectory in each family passes through the origin, and this trajectory, called the switching curve, must be chosen as the final path of motion if optimum response is to be achieved.[1,2] Without further discussion, it is clear that the optimum response for initial conditions on error is of the form shown in Fig. 4. For an initial condition at point A, the P arc through A never reaches the origin, therefore the N arc is chosen. When this N arc reaches the switching curve, the function f must pass through zero, allowing the motion to follow the switching curve into the origin. Bushaw has derived the optimum switching curve for initial conditions greater than X or Y, but this region of very large errors is not of interest here. The point is that the switching function must be zero along the switching curve, positive to the right of it and negative to the left of it. Thus the switching function f is determined, and, if we agree to ignore the interesting task of physically realizing this switching function exactly, the problem is solved.

Effect of a Reference Input

Referring to equation 1, it is evident that the optimization just described is valid only if the terms containing the reference input are zero. The exact contribution of a reference input is more clearly seen by considering the differential equation of the phase trajectories, equation 2.

$$\frac{de'}{de} = \frac{-2De' - e + r'' + 2Dr' + r - \operatorname{sgn} f}{e'} \qquad (2)$$

For any general reference input $r(\tau)$, the phase trajectories derived from equation 2 are functions of time τ as well as e and e', resulting in a family of phase trajectories that intersect each other in the phase plane. Thus a given point in the phase plane may correspond to any number of trajectories, depending upon the time τ.

This difficulty can be circumvented by representing system motion in extended phase space: i.e., phase space that has the three co-ordinates e, e', and τ. In this space system motion ultimately approaches equilibrium values that are curves governed by

$$e' = 0$$
$$e(\tau, f) = r''(\tau) + 2Dr'(\tau) + r(\tau) - \operatorname{sgn} f \qquad (3)$$

Considering the previous work in this light, phase-plane portraits are parallel projections of extended phase-space portraits.[6] In the example shown, an equilibrium point, such as P, is a projection of a horizontal line in the e, τ plane, and the logarithmic spiral is a spiral about this equilibrium line in extended phase space. A sample P trajectory in extended phase space is indicated in Fig. 5. In general, a P trajectory will spiral about a curve $P(\tau)$ in the e, τ plane, and at any given time τ_1, the equilibrium point in the phase plane is the projection on the e axis of $P(\tau_1)$. Therefore, if the P and N families are constructed in extended phase space, the optimum response problem can be studied for a general reference input $r(\tau)$. This procedure will be used to examine the implications of optimum response for step function and square-

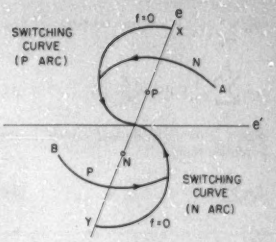


Fig. 4. Optimum response to initial errors

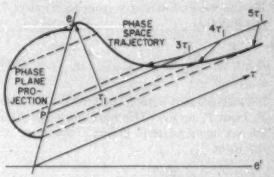


Fig. 5. P-trajectory in extended phase space

wave reference inputs to the linear, second-order underdamped system. These particular inputs are chosen because they are simple and because they represent transient and periodic inputs.

Effects of a Multiposition Discontinuous Device

If the discontinuous device has more than two necessary levels, there are more than two families of phase trajectories. Necessary levels are emphasized here because, if any level between the largest and the smallest can be eliminated, the error can be reduced to zero in shorter time by eliminating this level. A device with three necessary levels is a polar relay that takes a finite time to reverse. In such a case, the intermediate level is zero, and three families of phase trajectories exist.

If the general multiposition discontinuous device shown in Fig. 6 is represented by the function $\rho(f)$, where ρ is a constant whose value depends upon the value of the function f, the error differential equation may be written

$$e'' + 2De' + e = r'' + 2Dr' + r - \rho(f) \qquad (4)$$

This equation is an analytical statement of the problem and forms a basis for the study of optimum response of a control system having a multiposition discontinuous device. In a later section, certain general statements concerning optimum trajectories will be developed.

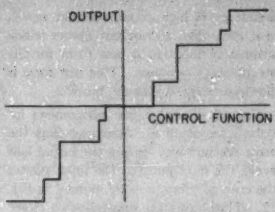


Fig. 6. A general amplitude sensitive discontinuous device

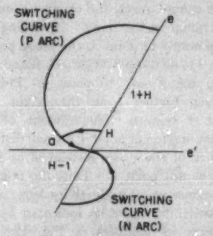


Fig. 7. Optimum switching function and step function response

from equation 4 and applied to the problem of optimum response using a relay with fixed time delay.

Optimum Response Using a 2-Position Switch

Step Function Response

If the reference input is a step function of magnitude H_1, the nondimensionalized input is H, where H must be less than unity (see Appendix). The error differential equation therefore becomes

$$e''+2De'+e=H-\operatorname{sgn} f \qquad (5)$$

It is immediately apparent that the phase plane trajectories are logarithmic spirals about equilibrium points, and are given by

$$e'=0$$
$$e=H-\operatorname{sgn} f \qquad (6)$$

For a positive step function input, the optimum switching curves are shown in Fig. 7. Since only two families exist, the argument for this case is exactly the same as before. The main difference in the result is that symmetry through the origin no longer exists, so that the P switching arc is larger than the N switching arc. Since the output cannot change instantaneously, the initial error is H, and the error response is HaO in Fig. 7.

A comparison of the output time response between an optimally switched relay system and a corresponding linear system is shown in Fig. 8 for $H=0.5$. Clearly, the relay system is far superior to the linear system. Depending upon the damping ratio D, the magnitude of the input step function H, and the allowable error criterion selected, the relay system reaches the allowable error in anywhere from 1/6 to 1/2 the time required by the linear system. Also it should be noted that the linear system is limited to reasonably larger values of D, while the relay system may use as small a value of D as can be physically achieved.

Fig. 8 also shows the locus of the initial arc (Ha in Fig. 7) of the response in the time domain for all values of H. All step function responses for the given value of D follow this locus until the relay switches. The analytic expression for this locus is the solution of the initial differential equation at the output shaft and is given by

$$C=1-\epsilon^{-D\tau}\left(\cos\beta\tau+\frac{D}{\beta}\sin\beta\tau\right)$$
$$\beta=\sqrt{1-D^2} \qquad (7)$$

In connection with this curve, it would be advantageous to have an expression for the time of switching. However, such an expression is a problem of the intersection of two spirals, and no analytic solution exists. Of course, a graphical solution could be obtained from the phase plane.

Referring again to Fig. 7, it is clear that optimum response can be attained only if the magnitude of the input step function is known, and the switching function is appropriately modified. If the effect of the input is ignored, a phase-plane sketch shows that zero error is reached because switching time lag produces a high-frequency chatter of the relay. Such a response is certainly not optimum. The fact that the optimum switching function depends upon the input should neither be surprising nor disconcerting since even linear optimum filter designs depend upon a knowledge of the input, especially if saturation is to be avoided.[7]

Square-Wave Reference Inputs

If the reference input is a square wave, the equilibrium values jump periodically from one location in the phase plane to another, as can be seen by inspection of the extended phase space diagram of Fig. 9 which shows the path of the P and N equilibrium points as time progresses. In the first interval ($0<\tau<\tau_s$), $P(\tau)$ is located at $e=P_1=1+H$ in the phase plane, and in the second interval, $P(\tau)$ is at $e=P_2=1-H$ in the phase plane. The N equilibrium point undergoes a similar series of jumps.

Since the equilibrium points are stationary over an interval, the phase trajectories remain logarithmic spirals about the appropriate point over this interval, and the previous argument for the step function is valid provided the half period τ_s is long enough. If τ_s is very short, the phase trajectory cannot intersect the switching function, and the response is essentially different. There-

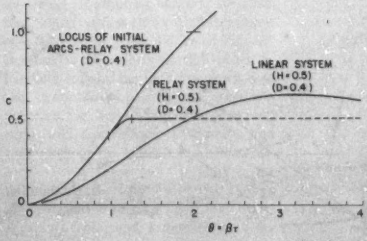


Fig. 8 (left). Step function response, relay versus linear system

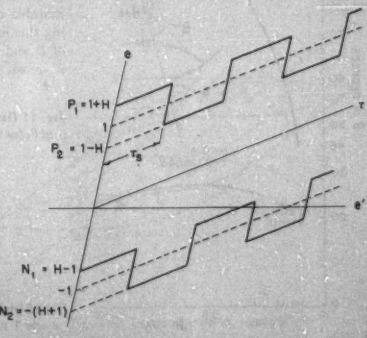


Fig. 9 (right). Extended phase space equilibrium values for square-wave reference input

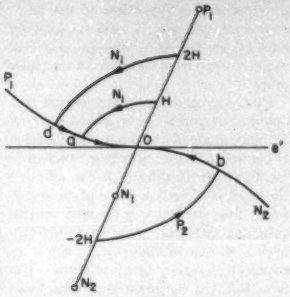


Fig. 10. Low-frequency mode of square-wave response

fore, two modes of operation present themselves; mode 1, the low frequency mode, and mode 2, the high frequency mode. It will be shown that these are the only two modes and that there is a marked transition between the two modes.

In mode, 1, τ_s is long enough for zero error to be reached, and the response may be found in the same manner as for the step function. A typical example is shown in Fig. 10. During the initial interval, the equilibrium points are P_1 and N_1; during the second interval, the equilibrium points are P_2 and N_2. Assuming the output to be zero at the time of application of the square wave, the initial error is H, and the switching curve, is the P_1 arc. The error then follows the N_1 arc to a, and from there to the origin along the P_1 arc. The error remains zero until the negative jump of $2H$ occurs in the reference input at time τ_s. Since the output cannot change instantaneously, the error jumps to $-2H$ follows a P_2 arc to the N_2 switching arc, point b, and then returns to the origin along the N_2 arc. At the next jump the initial error is $2H$, and the response is $2H$ to D to O, as indicated. From this time on the response repeats itself. It should be noted that only the P_1 and N_2 arcs enter into switching, and therefore the smaller P_2 and N_1 arcs may be omitted unless one is interested in perturbations superimposed on the square wave. This case is not considered here. It should be noted that after the initial transient in the first-half cycle, the response is in the steady state; i.e., the recurring transients are the same.

Any attempt to determine analytically the times of switching and zero error is immediately foiled since it requires the analytic solution to the intersection of spirals problem. These times can be measured from the phase plane plot.

Inspection of Fig. 10 reveals that the low-frequency mode ends when the following equation is satisfied.

$$\tau_{02} - \tau_s = \tau_s \tag{8}$$

where

τ_{02} = time of zero error during the second-half cycle

τ_s = half period of the square wave

Also from Fig. 10 it is evident that τ_{02} is a function of the amplitude of the input-square wave for a given load. For $D=0.4$ and $H=0.5$, the time $\tau_{02}-\tau_s$ is 1.80. Therefore, in this case, the lower limit on τ_s for low-frequency operation is 1.80. Suitable curves could be prepared showing this limiting value of τ_s for all values of D and H. One such curve for $D=0.4$ is shown in Fig. 11. If the square-wave frequency is high enough so that $\tau_{02}=2_s+\epsilon$, $\epsilon>0$, then a transient ensues which ultimately leads to a new form for the steady-state response. This new form is the high-frequency mode, mode 2.

Fig. 12 illustrates the phenomena involved for small ϵ. Assuming that the error reached zero during the initial half cycle, the next jump at the input causes the error to jump to $-2H$, point *1* in Fig. 12. The phase state then follows a P_2 arc to point *2*, where switching occurs, and the error decreases along the N_2 arc toward the origin. However, the phase state cannot reach the origin before the next jump in the input. Assuming this jump occurs at point *3*, the phase path is from *3* to *4*, along an N_1 arc to *5*, and then along the P_1 arc to point *6*. Point *6* is even further from the origin than point *3* because this segment of the motion started at point *4*, a small distance in back of the e axis instead of at the e axis, as did point *1*. Thus the point of the square-wave jump slides back along the switching curve, as indicated by the sequence of numbers in Fig. 12. For very small ϵ, the ultimate steady state is the path *19, 20, 21, 22, 19*. When this path has been reached, the switching function is still active, but zero error can never be attained.

In order to show that path *19, 20, 21, 22, 19* is indeed the final path, it is necessary to show that the phase state returns to its starting point if it starts anywhere on this trajectory. The geometrical configuration is that shown in Fig. 13, where only the left-half plane is considered because of the symmetry of the situation. Assuming the error starts at $2H$, the path is along the N_1 arc to point C, where the input and hence the error, jumps $2H$ to point D, as shown. The error then follows a P_2 arc to point F. If the distance OF is also $2H$, the path is repetitive, and the indicated path is the final path. The construction lines on Fig. 13 are used to effect this proof.

$\overline{AB}=2H$

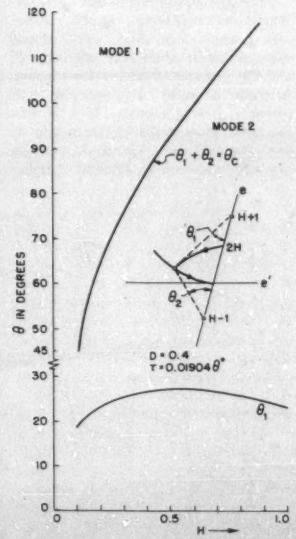


Fig. 11 (left). Limiting value of θ_s for mode 1 operation

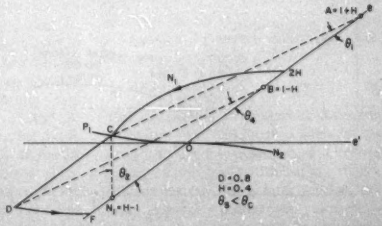


Fig. 12 (right). Transition from mode 1 to mode 2, small ϵ

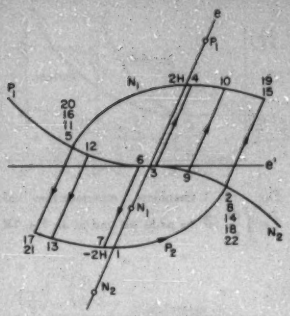


Fig. 13. Mode boundary

$\overline{CD}=2H$ by definition

Since $\overline{AB}$ and $\overline{CD}$ are parallel and equal, $\overline{AC}$ and $\overline{BD}$ are parallel and equal. Therefore $\overline{AC}=\overline{BD}$ and $\theta_1=\theta_4$

Since the spirals have the same damping constant and turn through the same angle, $\overline{AO}=\overline{BF}$, but $\overline{BF}=1-H+\overline{OF}$ and $\overline{AO}=1+H$. Therefore $\overline{OF}=2H$. Thus the indicated path is repetitive and represents the final path for small ϵ. That this path consumes the same amount of time as the limiting low frequency path as ϵ approaches zero is evident since $\theta_1=\theta_4$ and the net angle in each case is

$2(\theta_1+\theta_2)$ Mode 1, $2(\theta_4+\theta_3)$ Mode 2

This unique set of circumstances means that no hysteresis occurs in mode changing; i.e., for a given amplitude, the mode changes at a given value of τ_s whether τ_s is increasing or decreasing.

In order to determine the final trajectory in the high-frequency mode, consider the limit as the square-wave frequency approaches infinity. In this case the output cannot move, with the result that the error jumps back and forth from $+H$ to $-H$. As the frequency is decreased the final trajectory takes a form that satisfies all of the constraints of the system while maintaining the average error zero. A typical path for high-frequency operation is the path $ABCD$ of Fig. 14. Because of the algebraic difficulties in handling spirals, the high-frequency case was mapped graphically by choosing numbers for input amplitude, damping ratio, and square-wave frequency, and then plotting the resulting phase trajectory. From these plots a method for determining the location of the corners A, B, C, and D for any input amplitude, damping ratio, and frequency was deduced. The pattern that developed is shown in Fig. 15.

Consider any of the corners, say corner B. It was found that corner B followed the locus HE as frequency was decreased from infinity and that the locus HE is a spiral with the same damping constant as the phase trajectories. The center of this spiral is the point F which is found by constructing a stright line from the point E that makes an angle of $2\theta_s$ with EG and is H units long. The quantity θ_s is defined in terms of the square-wave half period by the following relation.

$$\theta_s=\beta\tau_s$$

$$\beta=\sqrt{1-D^2} \qquad (9)$$

Therefore the final path for mode 2 operation can be drawn when the frequency and amplitude of the square wave is known. This procedure eliminates the time-consuming process of tracing all the corners through their transient location, and for a given square-wave amplitude and system load, the complete response for any frequency input can be rapidly constructed.

The general result of this section is that optimum switching in the sense previously defined can be attained if the input square-wave frequency is low enough and if it is recognized that the optimum switching junction depends upon the input. For high-frequency square waves, the performance index chosen does not apply since the error is never allowed to reach zero. As a matter of fact, in the high-frequency case, the exact location of the switching function is of little significance; i.e., the switching function can vary over a wide range without affecting the steady-state response or the transient that precedes it. Since the frequency of mode changing can be calculated, the upper limit of optimum response can be definitely prescribed.

General Reference Inputs

If the reference input is an arbitrary time function, the optimum switching curves can be obtained from the trajectories in extended phase space. In general these curves will depend upon the input, but certain exceptions occur when the system differential equation has missing terms. In such cases, the optimum switching function may be the same for several different reference inputs. For example, if the system differential equation has no spring term, the initial condition and step function trajectories are the same, and no change in the optimum switching function is necessary.

In order to avoid the labor of plotting many extended phase-space trajectories, an arbitrary reference input may be approximated by step functions. If the input changes either very rapidly or very slowly, such a representation should yield a reasonable approximation to system optimum performance. However, medium changes in the input, when represented in this way, still lead to considerable labor in preparing an optimum switching function. It is clear from the square-wave discussion that optimum response in the sense defined does not apply to high-frequency periodic inputs.

Optimum Response Using a Multiposition Switch

General Criteria for Minimal Paths

The optimum switching function for an instantaneous 2-position switch driving an inertia load was obtained first by a heuristic argument.[1] Later, D. W. Bushaw proved rigorously that this result was indeed correct and, in addition, derived the optimum switching function for a general second-order load driven by an instantaneous 2-position switch with no reference input.[2] The broad outline of Bushaw's argument is used here to develop criteria for deriving optimum switching functions for a multiposition

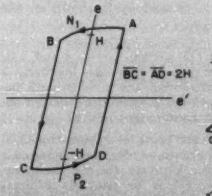


Fig. 14. Mode 2, high-frequency mode

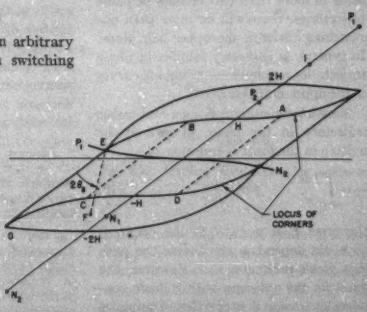


Fig. 15. Locus of corners, mode 2

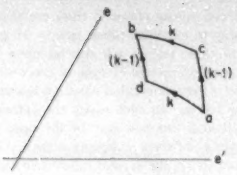


Fig. 16. Junction properties

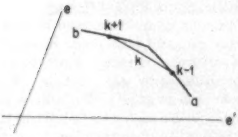


Fig. 17. Constraints on trajectory length

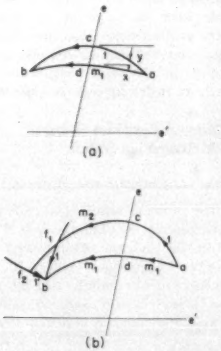


Fig. 18. Paths between two points, a and b, on the m trajectory

switch. Since the use of such a switch leads to more than two families of phase trajectories, there will be more than one switching curve in the upper half plane. In general, m necessary families of phase trajectories require $m-1$ switching curves in the upper-half plane.

Since a time interval along a continuous trajectory in the phase plane may be written in the form of equation 10,

$$\tau_{12}=\int_{e_1}^{e_2}\frac{de}{e'} \qquad (10)$$

the proof that a path to the origin is optimum depends upon finding the path that yields minimum τ_{12}. However, the quest for the optimum path is made considerably easier if properties of minimal paths between two phase states are established. A minimal path is defined as that path which requires the least time for the phase state to traverse from one point to another. If the two end points of the path lie in the same quadrant, the path with the largest e' for each e is minimal. If the two end points of the path lie in different quadrants, equation 10 is an improper integral that must be evaluated by a limiting process to determine the minimal path. The important item to note in the following development is that minimal paths are easily recognized by the character of their junctions, where a junction is a point at which the phase state passes from one family to another. Thus the following theorems on junctions allow determination of minimal paths.

Let there be m families of phase trajectories. If their slopes everywhere in the phase plane are known, the arcs can be numbered 1 through m, where 1 has the highest magnitude slope in quadrant I, and m has the lowest magnitude slope in quadrant I. If only paths that do not cross the e' axis are considered, the following theorems apply.

Theorem I. If $k\leq m$, a minimal path has junctions only of the form $(k-1)k$ in the right-half plane, and $k(k-1)$ in the left-half plane.

Proof: Suppose a $k(k-1)$ junction exists between two phase states a and b as shown at d in Fig. 16, then the $k(k-1)$ junction can be replaced by a path acb with a $(k-1)k$ junction that reaches b faster than adb. Therefore the theorem is proved.

Theorem II. At least $(m-2)$ constraints must be specified on the length of the m trajectories.

Proof: Suppose there were no constraint on the length of the kth trajectory, then the k arc can be removed by a $(k-1)(k+1)$ junction and the point b reached in a shorter time, as shown in Fig. 17. Thus, the theorem is proved.

Theorems I and II establish minimal paths in the same quadrant with the exception of the obvious case of the points a and b lying on the m trajectory. In such a case, the minimal path has no junctions. This statement is true even if the axis is crossed as is shown in the following proof.

Theorem III. If two points, a and b lie on an m trajectory that does not make a complete revolution, the minimal path between a and b has no junctions provided the family of m trajectories converge.

Proof: This proof is obvious for $1/e'$ always finite since no other path between a and b exists. However, the real usefulness of theorem III is to establish minimal paths when the axis is crossed, and in this case many paths exist between a and b. This can be seen by inspection of equation

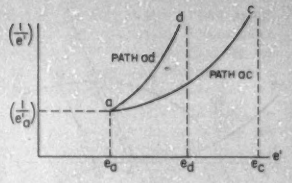


Fig. 19. Graphical representation of $\tau=\int\frac{de}{e'}$ for the paths ad and ac in Fig. 18

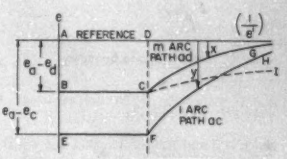


Fig. 20. Comparison of areas for paths ac and ad

10. When e' is positive, de must also be positive if positive time flow is to be maintained. Therefore all phase trajectories in the right-half plane point upward. Conversely, if e' is negative, de must also be negative, and all left-half plane trajectories point downward. Consequently, all phase trajectories that cross the e axis do so at zero slope, with negative slope in quadrant I and positive slope in quadrant II. This process is illustrated in Fig. 18, where it can be seen that the points a and b may be connected by positive time phase trajectories in many ways. The two ways shown represent the boundaries for all possible paths. Now it is necessary to prove that the m_1 path in Fig. 18 is minimal.

In order to accomplish this proof in general, that is for phase trajectories not necessarily spirals, consider the plots of $(1/e')$ versus e for the segments as shown in Fig. 19. In this plot, time intervals are areas between the curve of $1/e'$ and the e axis. In order to compare the areas ade_de_a and ace_ce_a, they are redrawn in Fig. 20, so that their asymptotes (e_ge and e_he) coincide. Although the path ad is shown as CG in Fig. 20, it may conceivably be CI. It will now be proved that its path is CG; i.e., it always lies between FH and the reference, by showing that the distance y is always greater than x for a given e'. Therefore path ad reaches the e axis in less time than path ac.

The distances x and y in the phase plane are shown in Fig. 18. Since the derivatives and one point on each phase curve are known, the distances x and y may be found by the integrations indicated in equations 11 and 12.

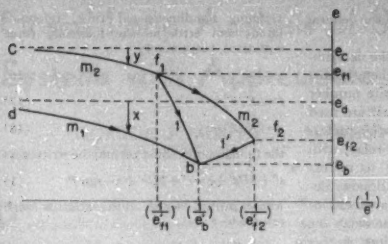


Fig. 21. Paths in e versus $1/e'$ plane for paths cf_1b, cf_2b, and db in the e versus e' plane

$$x = \int_{\epsilon_d}^{\epsilon_a} \left(\frac{de}{de'}\right)_m de' \tag{11}$$

$$y = \int_{\epsilon_c}^{\epsilon_a} \left(\frac{de}{de'}\right)_1 de' \tag{12}$$

But by definition

$$\left|\frac{de}{de'}\right|_1 > \left|\frac{de}{de'}\right|_m \text{ for } \epsilon' = \text{constant} \neq 0 \tag{13}$$

Since both slopes must be the same sign, y must be greater than x and path ad reaches the e axis before path ac.

To prove theorem III completely, it is now necessary to consider the arcs in the left-half plane. For this part, Figs. 18(A) and 18(B) must be considered separately. It is evident that the preceding argument can be applied directly to paths cb and db in Fig. 18(A) to prove the theorem for this configuration.

The proof for Fig. 18(B) proceeds in like manner. However, in this case it is necessary to distinguish between families of converging and diverging m trajectories because the theorem cannot be proved for diverging trajectories without specific knowledge of the functional relation between e and e' on the trajectory.

Considering converging or parallel m phase trajectories, the phase portrait is that shown in Fig. 18(B). From this figure, the plot of e versus $1/e'$ shown in Fig. 21 is obtained. For these m trajectories, the distance y is always greater than or equal to x for a given value of e'. Therefore, if Fig. 21 is redrawn so that the areas in question have a common infinite base line, the comparison becomes that shown in Fig. 22. From Fig. 22, it is evident that the area associated with the zero junction path m_1 is always less than that associated with the switched path. Since these areas are time intervals, theorem III is proved.

In the case of diverging trajectories, no general proof may be obtained. Using theorems I and III with the constraints required by theorem II, minimal paths may be constructed between any two points in the phase plane for any number of families of phase trajectories and for any form of trajectory, provided the m trajectories are converging. Since the argument depends only upon having a family of nonintersecting phase trajectories whose slopes are known, the theorems apply to piece-wise linear, nonlinear, or time varying trajectories. Since the optimum path is the minimal path from any point in the plane to the origin, optimum response may be determined as follows.

1. Given a set of phase trajectories, determine their slopes everywhere in the plane.
2. Apply theorems I, II, and III to construct minimal paths to the origin.
3. From these minimal paths select the optimum switching curves.

Optimum Response Using a Relay with Time Delay

All relays have some time delay τ_d in switching that usually depends upon the signal driving the relay. During this transition period, the relay load is driven by a function that depends upon the contact arcing characteristic. To simplify the analysis, the relay will be assigned constant time delay τ_d, and the relay output during transition will be assumed zero.[8] Three families of phase trajectories need to be considered; the P family, the N family, and O family. The O family spirals about the origin for the given load and is constrained to separate the P and N trajectories by a constant time τ_d. In the phase plane this constant time delay corresponds to a constant angle about the origin. Therefore the optimum switching curves and the optimum response are easily derived.

The resulting curves and the response for a disturbance input is shown in Fig. 23. It should be noted that if the initial condition on error lies between the actual switching curve and the switching demand curve, two switchings must take place before the response reaches the origin.

In response to a step-function reference-input of nondimensionalized value H, all

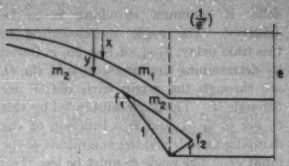


Fig. 22. Comparison of time areas, converging m arcs

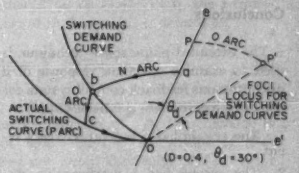


Fig. 23. Optimum disturbance response using a relay with time delay

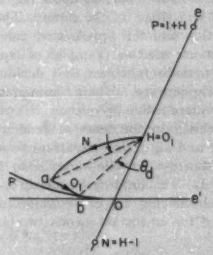


Fig. 24. Optimum step function response using a relay with time delay

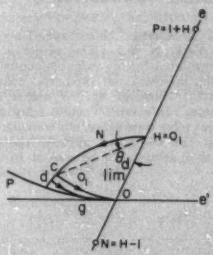


Fig. 25. Limiting time delay for optimum switching

three equilibrium points jump H units above their zero reference input location. Since the initial error is H, the optimum response is $HabO$ shown in Fig. 24. An interesting aspect of this type of response is that there is a limiting value of time

delay if optimum switching is to be achieved. The angle corresponding to this time delay, $\theta_{d\ lim}$ of Fig. 25, is found by determining the intersection of the O_1 arc through the origin with the N arc through H. The angle subtended by this O_1 arc is $\theta_{d\ lim}$ and is a function of the damping ratio D and the magnitude of the input H. The angle θ_d must be less than $\theta_{d\ lim}$ or optimum switching in the sense defined is impossible.

Conclusions

The extended phase-space technique is used to examine optimum response of a discontinuous feedback control system for the cases where these conditions apply:

1. The system has a general reference input.

2. The system has a multiposition discontinuous device.

It is shown that the optimum response characteristic depends upon the magnitude and shape of the input. The step function and the square-wave reference-input are used as examples of transient and periodic functions that demonstrate this dependence. There is an upper limit on square wave frequency if optimum switching, in the sense of reducing the error and all of its derivatives to zero in the shortest possible time, is to be attained. This limiting frequency is determined for an instantaneous two-position order-load as an illustration of the general method.

Criteria are developed for the determination of optimum response when the discontinuous device has any finite number of output levels. These criteria are used to derive the optimum response characteristics of a system having a relay with time delay between reversals, as an example of a three-level discontinuous device. It is shown that for a given input there may be limiting time delay for optimum response. The value of this limiting time delay is determined for the step function input.

Appendix

The fundamental differential equation relating the error e_0 and the input θ_i of the complete second-order system may be written

$$J\frac{d^2e_0}{dt^2}+B\frac{de_0}{dt}+Ke_0=J\frac{d^2\theta_i}{dt^2}+B\frac{d\theta_i}{dt}+K\theta_i-T\ \mathrm{sgn}\ f \quad (14)$$

The parameters J, B, and K are the inertia, viscous damping, and spring constant, respectively, on the output shaft, and T is the constant-torque output of the driving motor. Defining the ratio K/J as λ_0^2 and dividing through by T yields

$$\frac{K}{T\lambda_0^2}\frac{d^2e_0}{dt^2}+\frac{B}{T}\frac{de_0}{dt}+\frac{K}{T}e_0=\frac{K}{T\lambda_0^2}\frac{d^2\theta_i}{dt^2}+\frac{B}{T}\frac{d\theta_i}{dt}+\frac{K\theta_i}{T}-\mathrm{sgn}\ f \quad (15)$$

Defining nondimensional time, reference input, and error as the following three equations, respectively

$$\tau=\lambda_0 t \quad (16)$$

$$r=(K/T)\theta_i \quad (17)$$

$$e=(K/T)e_0 \quad (18)$$

the fundamental equation may be written as

$$e''+2De'+e=r''+2Dr'+r-\mathrm{sgn}\ f \quad (1)$$

where primes denote differentiation with respect to τ and

$$D=\frac{B}{2\sqrt{KJ}}$$

References

1. Nonlinear Techniques for Improving Servo Performance, D. McDonald. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 400–21.

2. Differential Equations with a Discontinuous Forcing Term, D. W. Bushaw. *Doctoral Dissertation*, Princeton University, Princeton, N. J., June, 1952.

3. An Investigation of Switching Criteria for Higher Order Contactor Servomechanisms, I. Bogner, L. F. Kazda. *AIEE Transactions*, vol. 73, pt. II, July 1954, pp. 118–26.

4. Introduction to Nonlinear Mechanics (book), N. Minorsky. J. W. Edwards, Ann Arbor, Mich.

5. Discontinuous Automatic Control, I. Flugge-Lotz. Princeton University Press, Princeton, N. J., 1953.

6. Optimum Switching in a Relay Servomechanism, F. W. Nesline, Jr. *Ph.D. Dissertation*, Yale University, New Haven, Conn., June, 1956.

7. Compensation of Feedback-Control Systems Subject to Saturation, G. C. Newton. *Journal*, Franklin Institute, Philadelphia, Pa., pt. I, Oct. 1952, pp. 281–96; pt. II, Nov. 1952, pp. 391–413.

Discussion

S. S. L. Chang (New York University, New York, N. Y.): The author is to be congratulated for his valuable contribution to the literature of predictor controlled systems. I should like to ask a simple question. In a multiposition relay system without relay time delay, nor transportation lag of the controlled system, the representative point is to be moved from one point in phase space to another in minimum time. Does the minimal path consist of the trajectories of the two outermost positions only? Or does it not?

F. W. Nesline, Jr.: The answer to Professor Chang's question is contained in theorem II of the paper which states that switches having no position constraints yield optimum response when only the outermost positions are used. Therefore an optimum path for a nonphysical (no constraint) system consists of the trajectories generated by two outermost positions only. However any physical relay has at least one transition region between the two outermost positions and is in effect a multiposition switch. The purpose of the second part of the paper is to show how the optimum response of a system containing such a multiposition switch can be determined and to show some of the consequences of this determination.

Closed-Loop Analysis of Sampled-Data Systems with Appreciable Pulse Width

GORDON J. MURPHY
ASSOCIATE MEMBER AIEE

HAROLD B. KENNEDY
NONMEMBER AIEE

DURING the past decade, interest in sampled-data systems has grown at a prodigious rate. In some of the early work[1,2] such systems were regarded as pulsed memory units which operate linearly on discrete data. In later work[3,4] the concept of impulse modulation was introduced, and the modern theory of sampled-data system analysis was established with the definition of the z-transformation.[4]

In the introduction of the concept of impulse modulation, a correspondence was established between the strength of an impulse function occurring at a time t_k and the magnitude at the time t_k of the signal being sampled. Thus, the set of discrete values of the sampled signal was converted to a train of impulses of appropriate strengths. The determination of the response of a linear system to a sampled signal was in this manner reduced to the determination of the response of the linear system to a modulated train of impulses. Numerous papers[5] published since the introduction of the impulse-modulation concept have been based thereon and have extended the work in various manners. However, in almost all cases the sampled version of a given signal continued to be regarded idealistically as a train of impulses of infinitesimal duration.

More recently, efforts have been made[6-8] to deal more realistically with the sampling problem by recognizing that in practice the duration of a pulse formed by sampling a function of time is not only not infinitesimal, but often appreciable. Farmanfarma[6,8] introduced the ρ-transform to allow consideration of pulse width, and Tou[7] introduced the τ-transform for the same purpose. These efforts led to the establishment of methods of analysis and synthesis of open-loop sampled-data systems with appreciable pulse width. Furthermore, both Farmanfarma and Tou have expressed the desirability of obtaining a method for the analysis and synthesis of closed-loop sampled-data systems with appreciable pulse width and have implied the existence of such a method. (Since this paper was completed, Farmanfarma has published[9] an exact method for the analysis and synthesis of closed-loop sampled-data systems with appreciable pulse width. However, the exact method is very tedious in application, and the approximations presented here continue to have considerable merit.)

The purpose of this paper is to present a method which allows the treatment of closed-loop sampled-data systems with pulse width ranging from zero to the sampling period. The technique presented here yields approximately the correct value of the actuating signal in a closed-loop system at instants in time separated by the sampling period; and, under certain conditions, with additional labor it yields approximately the correct value of the system response at all instants.

Analysis of Idealized Sampled-Data Systems

In the well-established use of the z-transformation in the study of sampled-data systems with infinitesimal pulse width, a periodic sampler is regarded as an impulse modulator in which a carrier in the form of a train of unit impulses is amplitude-modulated by the signal being sampled. The impulse train, denoted by $i(t)$, is expressed as

$$i(t) \triangleq \sum_{k=-\infty}^{\infty} u_0(t-kT) \tag{1}$$

where $u_0(t)$ denotes the unit impulse function with discontinuity at $t=0$ and T is the sampling period. If the function $f(t)$ being sampled, as illustrated in Fig. 1, is identically zero for $t<0$, then the output of the sampler is

$$f^*(t) = f(t)i(t) \tag{2}$$

$$= \sum_{k=0}^{\infty} f(kT)u_0(t-kT) \tag{3}$$

The Laplace transform of the sampled signal is

$$F^*(s) \triangleq \mathcal{L}f^*(t) \tag{4}$$

$$F^*(s) = \mathcal{L}[f(t)i(t)] \tag{5}$$

$$= \sum_{k=0}^{\infty} f(kT)e^{-kTs} \tag{6}$$

It is customary to simplify the right-hand side of equation 6 by introducing a new variable

$$z \triangleq e^{sT} \tag{7}$$

and to refer to the resulting expression as either $F^*(z)$ or $F(z)$. Thus, the z-transform of the signal $f(t)$ is defined as

$$F(z) = \sum_{k=0}^{\infty} f(kT)z^{-k} \tag{8}$$

Although the z-transform $F(z)$ is a unique function of $f(t)$, the converse is not true. That is, knowledge of $F(z)$ does not imply knowledge of $f(t)$ except at the sampling instants

$$t = kT, \quad k = 0, 1, 2, \ldots \tag{9}$$

Accordingly, one may not say that the inverse z-transform of $F(z)$ is $f(t)$. However, it may be said that it has the values of $f(0)$, $f(T)$, $f(2T)$, ..., at the instants $t=0, T, 2T, \ldots$, respectively, where $f(t)$ is any function of time the z-transform of which is the given $F(z)$. With this in mind, tables of z-transforms and inverse z-transforms can be compiled.[4]

If the sampler is followed by a linear system with transfer function $G(s)$, as shown in Fig. 2, the system response is

$$c(t) = \int_0^t g(\tau)r^*(t-\tau)d\tau \tag{10}$$

or, in terms of Laplace transforms,

$$C(s) = G(s)R^*(s) \tag{11}$$

At times it is convenient to introduce an imaginary sampler at the output of the system, as shown in Fig. 2. The Laplace transform of the sampled output then becomes of interest, and can be obtained by operating on both sides of equation 11 to obtain

$$C^*(s) = [G(s)R^*(s)]^* \tag{12}$$

It can be shown[10] that

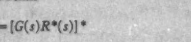


Fig. 1. An idealized sampler

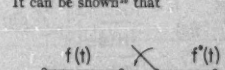


Fig. 2. An idealized open-loop sampled-data system

Paper 58-1303, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted June 30, 1958; made available for printing September 10, 1958.

GORDON J. MURPHY is with Northwestern University, Evanston, Ill., and HAROLD B. KENNEDY, formerly with Northwestern University, is now with Hughes Aircraft Company, Los Angeles, Calif.

$$[\overline{G(s)R^*(s)}]^* = G^*(s)R^*(s) \quad (13)$$

and hence that

$$C(z) = G(z)R(z) \quad (14)$$

The response at the sampling instants of a closed-loop sampled-data system can also be determined by application of equation 14. For example, for the system illustrated in Fig. 3(A),

$$E_1(z) = R(z) - B(z) \quad (15)$$

$$B(z) = E_1(z)\left\{[G(s)H(s)]^*\right\}\Big|_{e^{sT}=z} \quad (16)$$

and

$$C(z) = G(z)E_1(z) \quad (17)$$

For convenience it is customary to introduce a new symbol

$$GH(z) \triangleq \left\{[G(s)H(s)]^*\right\}\Big|_{e^{sT}=z} \quad (18)$$

with the result that equation 16 may be written

$$B(z) = GH(z)E_1(z) \quad (19)$$

Finally, simultaneous solution of equations 15, 17, and 19 yields the closed-loop relation

$$C(z) = \frac{G(z)}{1+GH(z)}R(z) \quad (20)$$

For a system such as that illustrated in Fig. 3(B), the corresponding relation is

$$C(z) = \frac{G(z)}{1+G(z)H(z)}R(z) \quad (21)$$

and other similar relations can be derived[4] for other closed-loop systems.

Analysis of Sampled-Data Systems with Appreciable Pulse Width

If the duration of the sampling pulse is not negligible, a slightly different approach is used. The impulse train $i(t)$ is replaced by a train of pulses of unit height and repetition period T, and the width of each pulse is denoted by h. Such a pulse train, illustrated in Fig. 4, is denoted[9] by $u_p(t)$ and may evidently be written

$$u_p(t) = \sum_{n=-\infty}^{\infty}[u_{-1}(t-nT) - u_{-1}(t-nT-h)] \quad (22)$$

where $u_{-1}(t)$ is the unit step function with discontinuity at $t=0$.

If the function being sampled, as shown in Fig. 5, is $f(t)$, then the output of the sampler is

$$f_p(t) = f(t)u_p(t) \quad (23)$$

The Laplace transform of the sampled signal, referred to as the p-transform[8] of $f(t)$, is

$$F_p(s) = \mathcal{L}[f(t)u_p(t)] \quad (24)$$

$$= \int_0^h e^{-st}f(t)dt + \int_T^{T+h} e^{-st}f(t)dt + \int_{2T}^{2T+h} e^{-st}f(t)dt + \ldots \quad (25)$$

$$= \sum_{k=0}^{\infty}\int_{kT}^{kT+h} e^{-st}f(t)dt \quad (26)$$

Tables[8] of p-transforms can be compiled and used to facilitate the analysis and synthesis of sampled-data systems with appreciable pulse width. However, it must be remembered that, as is true of z-transforms, there is no unique inverse p-transform. The inverse p-transform of a given $F_p(s)$ is defined only for $nT \leq t < nT + h$, $n = 0, 1, 2, \ldots$; but for values of h that are an appreciable percentage of T, the inverse p-transform conveys considerably more information concerning the time function than does the inverse z-transform. In fact, in such instances the problem of intersample ripple[11] is reduced to one of minor significance.

If the generalized sampler is followed by a linear system with transfer function $G(s)$, as shown in Fig. 6, then the system response is

$$c(t) = \mathcal{L}^{-1}G(s)R_p(s)$$

$$= \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} e^{st}G(s)R_p(s)ds \quad (27)$$

Thus, the response of the system for all $t > 0$ can be determined by contour integration. Alternatively, the response of such a system may be written

$$c(t) = \int_0^t r_p(\tau)g(t-\tau)d\tau \quad (28)$$

$$= \sum_{k=0}^{\infty}\int_{kT}^{kT+h} r(\tau)g(t-\tau)d\tau \quad (29)$$

The response of a closed-loop sampled-data system with appreciable pulse width is considerably more difficult to determine. One method proposed here consists of "decomposing" the sampler with pulse width h into a number of samplers of pulse width $\tau < h$. Each of the resulting samplers can then be regarded as approximately equivalent to an ideal sampler, and the wealth of z-transform theory can be brought to bear on the problem.

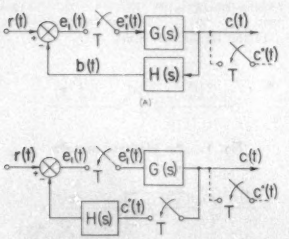


Fig. 3. Two idealized closed-loop sampled-data systems

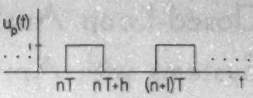


Fig. 4. Pulse train

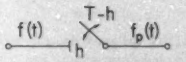


Fig. 5. A generalized sampler

The method can be developed with the aid of Figs. 7 and 8. The system illustrated in Fig. 7 is a general single-loop "error"-sampled system in which a periodic sampler with pulse width h and period T is employed. The input to the sampler can be approximated by

$$\bar{e}_1(t) = e_1(0)[u_{-1}(t) - u_{-1}(t-\tau)] + e_1(\tau)[u_{-1}(t-\tau) - u_{-1}(t-2\tau)] + e_1(2\tau)[u_{-1}(t-2\tau) - u_{-1}(t-3\tau)] + \ldots \quad (30)$$

$$= \sum_{k=0}^{\infty} e_1(k\tau)[u_{-1}(t-k\tau) - u_{-1}(t-k\tau-\tau)] \quad (31)$$

where τ is a small positive quantity. When $e_1(t)$ is thus approximated, with $\tau = h/p$, where p is an integer, the sampled actuating signal is

$$e_{1p}(t) \approx \sum_{k=0}^{p-1} e_1(k\tau)[u_{-1}(t-k\tau) - u_{-1}(t-k\tau-\tau)] + \sum_{k=0}^{p-1} e_1(T+k\tau)\times[u_{-1}(t-T-k\tau) - u_{-1}(t-T-k\tau-\tau)] + \ldots \quad (32)$$

$$= \sum_{n=0}^{\infty}\sum_{k=0}^{p-1} e_1(nT+k\tau)[u_{-1}(t-nT-k\tau) - u_{-1}(t-nT-k\tau-\tau)] \quad (33)$$

Rearranging the terms on the right-hand side of equation 33 gives

$$e_{1p}(t) \approx \sum_{k=0}^{p-1}\sum_{n=0}^{\infty} e_1(nT+k\tau)[u_{-1}(t-nT-k\tau) - u_{-1}(t-nT-k\tau-\tau)] \quad (34)$$

$$= \sum_{k=0}^{p-1}\tau\sum_{n=0}^{\infty} e_1(nT+k\tau)\times\left[\frac{u_{-1}(t-nT-k\tau) - u_{-1}(t-nT-k\tau-\tau)}{\tau}\right] \quad (35)$$

The term on the right-hand side of equation 35 obtained for $k=0$ is

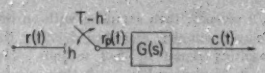


Fig. 6. A generalized open-loop sampled-data system

$$e_{11}(t) = \tau \sum_{n=0}^{\infty} e_1(nT) \times \left[\frac{u_{-1}(t-nT-\tau) - u_{-1}(t-nT-\tau)}{\tau}\right] \quad (36)$$

$$\approx \tau \sum_{n=0}^{\infty} e_1(nT)\, u_0(t-nT) \quad (37)$$

$$= \tau e_1^*(t) \quad (38)$$

The term on the right-hand side of equation 35 obtained for $k=1$ is

$$e_{12}(t) \approx \tau \sum_{n=0}^{\infty} e_1(nT+\tau) \times \left[\frac{u_{-1}(t-nT-\tau) - u_{-1}(t-nT-2\tau)}{\tau}\right] \quad (39)$$

$$\approx \tau \sum_{n=0}^{\infty} e_1(nT+\tau)\, u_0(t-nT-\tau) \quad (40)$$

$$= \tau_\tau e_1^*(t) \quad (41)$$

where ${}_\tau f^*(t)$ is defined in general as the signal obtained by sampling the function $f(t)$ by means of an ideal periodic sampler with period T the actuation of which has been delayed by τ seconds. (That is, the sampling switch closes at $t=\tau,\ T+\tau,\ 2T+\tau,\ \ldots$, instead of at $t=0,\ T,\ 2T,\ \ldots$.)

The terms on the right-hand side of equation 35 obtained for $k=2,\ 3,\ \ldots,\ p-1$ may be interpreted in a similar manner, with the result that equation 35 may be rewritten

$$e_{1p}(t) \approx \sum_{k=0}^{p-1} \tau_{\ k\tau} e_1^*(t) \quad (42)$$

Thus, it is seen that the effect of the single periodic sampler with pulse width h and period T can be approximated by p ideal periodic samplers with period T, with their actions appropriately delayed, operating in parallel.

Since

$${}_{k\tau}f^*(t) = \sum_{n=0}^{\infty} f(nT+k\tau)\, u_0(t-nT-k\tau) \quad (43)$$

$$= f(t)i(t-k\tau) \quad (44)$$

$$= {}_{k\tau}f^*(t-k\tau) \quad (45)$$

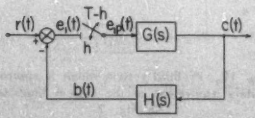


Fig. 7 (left). A closed-loop sampled-data system with appreciable pulse width

Fig. 8 (right). A system which is approximately equivalent to the system illustrated in Fig. 7

where

$${}_{k\tau}f^*(t) \triangleq f(t+k\tau)i(t) \quad (46)$$

it follows that the effect of an ideal periodic sampler with its actuation delayed by $k\tau$ seconds is identical to that of an ideal periodic sampler with the same period and no delay in actuation, provided that the signal to be sampled is advanced in time by $k\tau$ seconds before being operated on by the latter sampler and that the output of the latter sampler is retarded in time by $k\tau$ seconds. Accordingly, the system illustrated in Fig. 7 is approximately equivalent to the system shown in Fig. 8.

It is evident in Fig. 8 that

$$E_1(s) = R(s) - B(s) \quad (47)$$

and

$$B(s) = G(s)H(s)E_{1p}(s) \quad (48)$$

$$= \tau G(s)H(s) \sum_{k=0}^{p-1} e^{-k\tau s} E_{1(k)}^*(s) \quad (49)$$

where

$$E_{1(k)}(s) \triangleq e^{k\tau s} E_1(s) \quad (50)$$

Now let τ be such that in addition to satisfying the relation

$$\tau = \frac{h}{p} \quad (51)$$

it also satisfies the relation

$$\tau = \frac{T}{q} \quad (52)$$

where both p and q are integers. (Although equations 51 and 52 can be exactly satisfied simultaneously only if h/T is a rational number, if h/T is irrational, equations 51 and 52 can be satisfied to as good an approximation as desired, by suitable choice of p and q.) Then let

$$z_q \triangleq e^{sT/q} \quad (53)$$

Taking the Z_q transform of both sides of equations 47 and 49 yields

$$E_1(z_q) = R(z_q) - B(z_q) \quad (54)$$

and

$$B(z_q) = \tau Z_q\left[G(s)H(s)\sum_{k=0}^{p-1} e^{-(k\tau/q)s} E_{1(k)}(s)\right] \quad (55)$$

Since $e^{-(k\tau/q)s}E_{1(k)}^*(s)$ is periodic in s with period $j2\pi/T$ for $k=0$ and period $j2\pi q/kT$ for $k=1,\ 2,\ \ldots,\ p-1<q$, it follows that equation 55 may be written

$$B(z_q) = \left[\tau \sum_{k=0}^{p-1} z_q^{-k} E_{1(k)}(s)\right] GH(z_q) \quad (56)$$

and combining equations 54 and 56 yields

$$E_1(z_q) + \left[\tau \sum_{k=0}^{p-1} z_q^{-k} E_{1(k)}(s)\right] GH(z_q) = R(z_q) \quad (57)$$

It can be shown[15] that

$$E_{1(k)}(s) \equiv \frac{1}{q} z_q^k \sum_{n=1}^{q} e^{j2\pi k(q-n)/q} E_1(z_q e^{j2\pi(q-n)/q}) \quad (58)$$

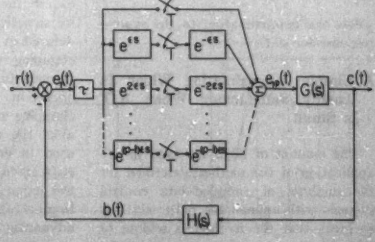
Substituting equation 58 into equation 57 gives

$$E_1(z_q) + \frac{\tau}{q}\left\{\sum_{k=0}^{p-1} e^{j2\pi k} \sum_{n=1}^{q} e^{-j2\pi kn/q} \times E_1[z_q e^{-j2\pi(q-n)/q}]\right\} GH(z_q) = R(z_q) \quad (59)$$

Replacing z_q by $z_q e^{j2\pi m/q}$, with $m=1,\ 2,\ \ldots,\ q-1$, in equation 59 yields $q-1$ additional independent equations of similar form. Thus, one obtains the q independent equations

$$E_1(z_q e^{j2\pi m/q}) + \frac{\tau}{q} \sum_{k=0}^{p-1} e^{j2\pi k} \sum_{n=1}^{q} e^{-j2\pi kn/q} \times E_1[z_q e^{j2\pi(m+q-n)/q}] GH(z_q) = R(z_q e^{j2\pi m/q}),\ m=0,\ 1,\ 2,\ \ldots,\ q-1 \quad (60)$$

in the q unknowns $E_1(z_q e^{j2\pi m/q})$, $m=0,\ 1,\ 2,\ \ldots,\ q-1$.



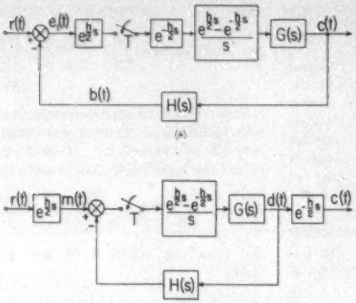


Fig. 9. A second system which is approximately equivalent to the system illustrated in Fig. 7

This set of q independent linear equations can be solved for the q unknowns (by use of matrices, for example), and then $E_{1p}(s)$ can be obtained by linearly combining the q solutions as suggested in equation 49. That is,

$$E_{1p}(s) \approx \tau \sum_{k=0}^{p-1} e^{-k\tau s} E_{1(k)}{}^*(s) \tag{61}$$

Substituting for $E_{1(k)}{}^*(s)$ the identity implied in equation 58 then yields the desired solution

$$E_{1p}(s) \approx \frac{\tau}{q} \sum_{k=0}^{p-1} \sum_{n=1}^{q} e^{j2\pi k(q-n)/q} \times E_1[z_q e^{j2\pi(q-n)/q}]\,|_{z_q = e^{sT/q}} \tag{62}$$

The approximate value of the actuating signal at each instant at which the sampling switch is in its closed position is

$$e_{1p}(t) = \mathcal{L}^{-1} E_{1p}(s) \tag{63}$$

$$\approx \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e^{st} \left\{ \frac{\tau}{q} \sum_{k=0}^{p-1} \sum_{n=0}^{q} e^{j2\pi k(q-n)/q} \times E_1[z_q e^{j2\pi(q-n)/q}]\,|_{z_q = e^{sT/q}} \right\} ds \tag{64}$$

Finally, substituting equation 62 into the expression

$$c(t) = \mathcal{L}^{-1} G(s) E_{1p}(s) \tag{65}$$

yields the approximation to the system response for all $t > 0$.

A Simple Approximation Which Is Usually Satisfactory When h/T Is Small

The amount of labor involved in the application of the method described for the analysis of sampled-data control systems with appreciable pulse width is so great that the method is seldom of practical value. For this reason a poorer approximation to the sampled signal which leads to a simplified analysis and yet yields results of satisfactory accuracy is highly desirable. One such approximation is described in the following.

An approximation to the input to the sampler in Fig. 7 which, although it is less accurate than that given in equation 30, is often satisfactory, is

$$\begin{aligned} e_{1'}(t) = {} & e_1\left(\frac{h}{2}\right)[u_{-1}(t) - u_{-1}(t-h)] + \\ & e_1(t)[u_{-1}(t-h) - u_{-1}(t-T)] + \\ & e_1\left(T+\frac{h}{2}\right)[u_{-1}(t-T) - u_{-1}(t-T-h)] + e_1(t)[u_{-1}(t-T-h) - u_{-1}(t-2T)] + e_1\left(2T+\frac{h}{2}\right)[u_{-1}(t-2T) - u_{-1}(t-2T-h)] + e_1(t)[u_{-1}(t-2T-h) - u_{-1}(t-3T)] + \ldots \end{aligned} \tag{66}$$

When $e_1(t)$ is thus approximated, the sampled actuating signal is

$$e_{1p}(t) \approx \sum_{k=0}^{\infty} e_1\left(kT+\frac{h}{2}\right) \times [u_{-1}(t-kT) - u_{-1}(t-kT-h)] \tag{67}$$

The right-hand side of equation 67 is exactly the signal that would be obtained by sampling $e_1(t)$ with an ideal sampler delayed in operation by $h/2$ seconds and clamping the resulting signal for $h/2$ seconds before and after the sampling operation. While it is true that such clamping action is not physically obtainable, the mathematical model may be used to describe the approximation to $e_{1p}(t)$ given in equation 67. Furthermore, the action of a sampler delayed in operation by $h/2$ seconds can be duplicated by advancing the signal to be sampled by $h/2$ seconds, then sampling with an undelayed sampler, and then delaying the resulting signal by $h/2$ seconds. The block diagram in Fig. 7 can, therefore, be replaced with the block diagram in Fig. 9(A) whenever the right-hand side of equation 66 gives a satisfactory approximation to $e_1(t)$. Finally, the analysis of the system is simplified by redrawing the block diagram in the entirely equivalent form shown in Fig. 9(B).

By the conventional use of z-transforms it can easily be shown that, for the system illustrated in Fig. 9(B),

$$\frac{D(z)}{M(z)} = \frac{G_1(z)}{1 + G_1 H(z)} \tag{68}$$

where

$$G_1(z) \triangleq \frac{e^{(h/2)s} - e^{-(h/2)s}}{s} G(s) \tag{69}$$

It follows that

$$D(z) = \frac{G_1(z)}{1 + G_1 H(z)} Z\left[\tau\left(t+\frac{h}{2}\right)\right] \tag{70}$$

Since

$$c(t) = d\left(t - \frac{h}{2}\right) \tag{71}$$

$$c\left(kT + \frac{h}{2}\right) = d(kT),\ k = 0, 1, 2, \ldots \tag{72}$$

can be obtained directly from $z^{-1}D(z)$.

The determination of the pulse transfer functions $G_1(z)$ and $G_1H(z)$ is complicated by the presence of the factor $e^{(h/2)s} - e^{-(h/2)s}$

However, these pulse transfer functions can be evaluated either by the use of the contour integral[11]

$$F^*(z) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{F(p)}{1 - e^{-T(z-p)}}\, dp \tag{73}$$

where c is less than the real part of s and greater than the real part of every pole of $F(p)$, or by the use of special tables.[12]

A Simple Approximation Which Is Usually Satisfactory When h/T Is Large

Although it is to be expected that the method proposed herein for analyzing

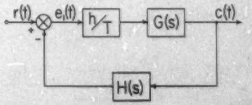


Fig. 10. A third system which is approximately equivalent to the system illustrated in Fig. 7

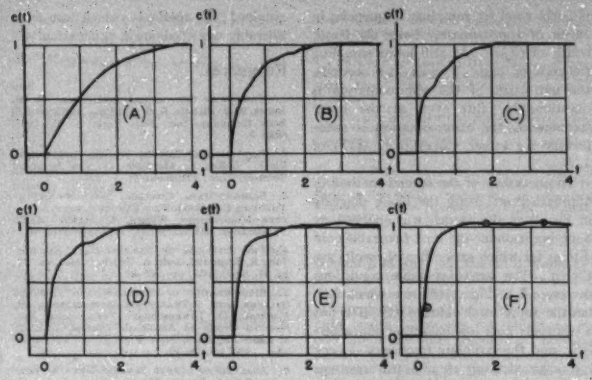


Fig. 11. Response to a unit step input of a system of the type illustrated in Fig. 7, with $G(s)=29.44(s+1)/s(s+0.8)(s+10)$, $H(s)=1$, and $h/T=0.25$

A—Approximation based on equation 77
B—$T=0.25$
C—$T=0.33$
D—$T=0.5$
E—$T=1.0$
F—$T=1.6$

sampled-data systems with finite pulse width will be satisfactory when h/T is small and the sampling frequency is not excessively low for a given $G(s)$ and $H(s)$, it is evident that unless the sampling frequency is extremely high, the same approximation will not be very good for large values of $h/T<1$. A different method of approximation, which is even simpler to employ and which yields results of greater accuracy for the larger values of h/T, is described in the following.

The periodic pulse train described by the right-hand side of equation 22 can be described alternatively by the Fourier series

$$u_p(t)=\frac{1}{T}\sum_{n=-\infty}^{\infty}\frac{1-e^{-jn\Omega h}}{jn\Omega}e^{jn\Omega t} \qquad (74)$$

Substituting equation 74 into 23 and carrying out term by term the indicated multiplication yields

$$f_p(t)=\frac{1}{T}\sum_{n=-\infty}^{\infty}\frac{1-e^{-jn\Omega h}}{jn\Omega}e^{jn\Omega t}f(t) \qquad (75)$$

Taking the Laplace transform of both sides of equation 75 and changing the order of summation and integration then yields

$$F_p(s)=\frac{1}{T}\sum_{n=-\infty}^{\infty}\frac{1-e^{-jn\Omega h}}{jn\Omega}F(s-jn\Omega) \qquad (76)$$

by application of the Complex Translation Theorem.

The maximum value of $|1-e^{-jn\Omega h}|$ is 2, and $|jn\Omega T|=2\pi n$. Therefore, if $|F_p(s-jn\Omega)|$ is a monotonically decreasing function of n, no term in the series on the right-hand side of equation 76 for $n\neq 0$ can have a magnitude that exceeds $1/\pi$ times the magnitude of the term for $n=0$. Furthermore, for a given $F(s)$, as $h/T\to 1$ the maximum value of each term in the series except the term for $n=0$ decreases. Therefore, as $h/T\to 1$, an increasingly better approximation is obtained by the use of only the term for $n=0$ on the right-hand side of equation 76, provided that $|F(s-jn\Omega)|$ decreases monotonically as n increases. This condition is satisfied at real frequencies less than $\Omega/2$ if the amplitude spectrum of $F(j\omega)$ is a monotonically decreasing function of ω on $0<\omega<\Omega/2$ and if $|F(j\omega)|<|F(j\Omega/2)|$ for $\omega>\Omega/2$. Furthermore, these conditions on the amplitude spectrum of $F(j\omega)$ are often satisfied in practice.

It follows from the foregoing that for $h/T\to 1$ a good approximation to $F_p(s)$ is often afforded by

$$F_{p0}(s)=\frac{h}{T}F(s) \qquad (77)$$

In other words, for $h/T\to 1$ the effect of the sampler is primarily to introduce an attenuation. A block diagram, drawn on this basis, of a system which for $h/T\to 1$ is approximately equivalent to the system illustrated in Fig. 7, is shown in Fig. 10. The value of this approximation is evident: the pulsed system behaves approximately as a continuous system, and in many instances the use of the block diagram in Fig. 10 for analysis and synthesis will yield results of sufficient accuracy for practical purposes.

There are some important limitations, however, on the use of the continuous-system approximation to the sampled-data system with appreciable pulse width as well as on the use of the approximation based on equation 66. Probably the major limitation in both cases is the failure of the approximation to account for the differences in response brought about by changing the instant during the sampling cycle at which the input is actually impressed upon the system. For design (or at least in the early phases of design), this is not a serious disadvantage. In fact, the conventional method of treating ideal sampled-data systems by the use of the z-transformation is similarly limited. And the analysis by the use of the approximations of sampled-data systems with appreciable pulse width can be satisfactorily effected by a modification of the techniques outlined. The description of this modification is too lengthy for inclusion here and hence is scheduled for presentation in another paper.

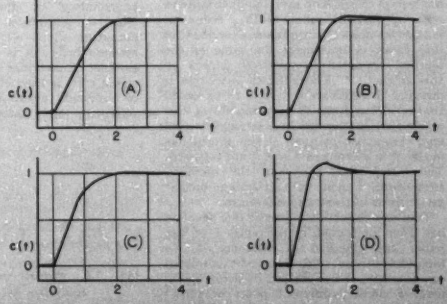


Fig. 12. Response to a unit step input of a system of the type illustrated in Fig. 7 with $G(s)=100/s(s+5)^2$, $H(s)=1$, and $h/T=0.25$

A—Approximation based on equation 77
B—$T=0.25$
C—$T=0.5$
D—$T=1.0$

Experimental Results

In an effort to determine the value of the approximations suggested in the foregoing, two typical sampled-data control systems with appreciable pulse width were simulated on an analog computer. The systems studied are both of the type illustrated in Fig. 7, with $H(s)=1$, and in both systems $h/T=0.25$. In the first of these systems to be discussed,

$$G(s)=\frac{29.44(s+1)}{s(s+0.8)(s+10)} \tag{78}$$

and in the second,

$$G(s)=\frac{100}{s(s+5)^2} \tag{79}$$

The results obtained are presented in Figs. 11 and 12.

In all cases the input to the system is a unit step function with its discontinuity ($t=0$) coincident with the initiation of a sampling pulse.

The curves in Figs. 11(A) and 12(A) were obtained by replacing the sampler with an attenuator, according to the approximation based on equation 77. The curves in Figs. 11(B), (C), (D), (E), and (F), and Figs. 12(B) and (C) were obtained by the accurate simulation of the systems under consideration, using an actual sampler with pulse width of $h=0.25T$. The sampling period in each case is identified in the caption.

In spite of the fact that the value of h/T in the systems simulated is only 0.25, the approximation to the actual response of the sampled-data system afforded by the continuous-system approximation [in Figs. 11(A) and 12(A)] is fairly good for sampling frequencies in excess of approximately twice the bandwidth of $G(j\omega)$. At still lower sampling frequencies (e.g., 1 cycle per second), the usefulness of the approximation is questionable. But even at the lower frequencies, the approximation is satisfactory for larger values (e.g., 0.75) of h/T.

The accuracy of the second method of approximation, using the block diagram in Fig. 9 as the model, was checked by applying it under the least favorable conditions for which experimental results are shown. The least favorable conditions correspond to Fig. 11(F), for which h/T has the appreciable value of 0.25, $G(s)$ has a relatively poor low-pass filter characteristic, and the sampling frequency is very low (approximately equal to the crossover frequency) for the given $G(s)$. The results obtained are indicated by the encircled points superimposed on the graph in Fig. 11(F).

Conclusions

The response of a closed-loop sampled-data system with finite pulse width can, with reasonable limitations on the nature of the system, be predicted with engineering accuracy by the use of the fairly simple techniques described here. Although the results obtained in this manner may not be sufficiently accurate to be used as the ultimate description of system response, they are at least of considerable value early in the design of a system. They have value also as a relatively easily obtained check on results, believed to be more accurate, obtained by methods which are considerably more tedious in application.

References

1. Theory of Servomechanisms (book), H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, Inc., New York, N. Y., 1947. chap. 5.

2. Fundamental Theory of Servomechanisms (book), LeRoy A. MacColl. D. Van Nostrand Company, Inc., Princeton, N. J., 1945.

3. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, William K. Linvill. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1779–88.

4. The Analysis of Sampled-Data Systems, John R. Ragazzini, Lotfi A. Zadeh. *Ibid.*, vol. 71, pt. II, Nov. 1952, pp. 225–32.

5. Bibliography of Sampled-Data Control Systems and Z-Transform Applications, H. Freeman, O. Lowenschuss. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. AC-4, Mar. 1958, pp. 28–30.

6. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, G. Farmanfarma. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 808–19.

7. Analysis of Sampled-Data Control Systems with Finite Sampling Duration, Julius Tou. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 13, 1957, pp. 561–73.

8. Analysis of Multiple Sampler Systems With Finite Pulse Width, Open Loop, G. Farmanfarma. *AIEE Transactions*, vol. 77, pt. II, Mar. 1958, pp. 20–28.

9. General Analysis and Stability Study of Finite Pulsed Feedback Systems, G. Farmanfarma. *Ibid.*, July, pp. 148–52.

10. Automatic Feedback Control System Synthesis (book), John G. Truxal. McGraw-Hill Book Company, Inc., 1955.

11. Analysis and Synthesis of Sampled-Data Control Systems, Eliahu I. Jury. *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

12. Input-Output Analysis of Multirate Feedback Systems, G. M. Kranc. *Transactions* Professional Group on Automatic Control, Institute of Radio Engineers, vol. AC-3, Nov. 1957, pp. 21–28.

13. Sampled-Data Processing Techniques for Feedback Control Systems, Arthur R. Bergen, John R. Ragazzini. *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 236–47.

Discussion

Julius Tou (Purdue University, Lafayette, Ind.): This paper presents an interesting analysis of closed-loop sampled-data systems with appreciable pulse width. However, it appears that the approach described in this paper is essentially the same as the method proposed in previous papers on this subject.[1-4] The authors' attention is directed to reference 3. It is to be noted that a closed form of equation 62 of this paper may be derived in a straightforward and systematic manner, if the procedures outlined in references 3 and 4 are followed. The z-transform $C(z)$ and the modified z-transform $C(z, m)$ of the system output are derived in these two references.

The approximate analysis of this problem has been presented in reference 7 of the paper and detailed in reference 4. In fact, the authors' analysis may be simplified considerably, if the problem is investigated from a slightly different point of view.[4] As can readily be seen, if the pulse width is narrow, the actual pulses may be approximated by equivalent flat-topped pulses as shown in Fig. 13, and the sampled-data system of Fig. 7 may be represented approximately by the block diagram of Fig. 14. Physically, this implies that the actuating signal $e(t)$ is advanced by $h/2$ seconds before it is sampled by the ideal sampler S, and that the sampled actuating signal $e^*(t)$ is then clamped for an interval of h seconds, where h is the pulse width in seconds. It is noted that $(1-\epsilon^{-hs})/s$ is the transfer function of a partially-clamped zero-order hold.[1] Thus, the signal applied to the controlled system $G(s)$ is a train of equivalent flat-topped pulses. Clearly, the analysis of the equivalent system shown in Fig. 14 is simple and straightforward, as demonstrated in the following paragraphs.

Making use of equation 18 of reference 3, it is obtained that the z-transform of the system output is

$$C(z)=\frac{Z\{(1-\epsilon^{-hs})G_s(s)\}\,Z\{\epsilon^{hs/2}R(s)\}}{1+Z\{\epsilon^{hs/2}(1-\epsilon^{-hs})G_s(s)H(s)\}} \tag{80}$$

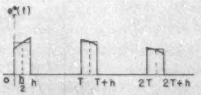


Fig. 13. Actual pulses represented by equivalent flat-topped pulses

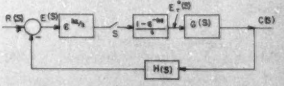


Fig. 14. Equivalent block diagram describing the system of Fig. 7

in which $G_s(s)=G(s)/s$. It has been shown that[4]

$$Z\{e^{Ms}G(s)\}=zG(z,m)\big|_{m=h}=zG(z,h) \quad (81)$$

and

$$Z\{e^{-Ms}G(s)\}=G(z,m)\big|_{m=1-h}=G(z,1-h) \quad (82)$$

where $G(z,m)$ is the modified z-transform associated with $G(s)$. Thus, equation 80 may be written as

$$C(z)=\frac{z[G_s(z)-G_s(z,1-h)]R(z,h/2)}{1+zG_sH(z,h/2)-G_sH(z,1-h/2)} \quad (83)$$

where $G_sH(z,h/2)$ is given by the modified z-transform associated with $G_s(s)H(s)$ evaluated at $m=h/2$. The modified z-transform $C(z,m)$ of the system output can also be derived in a similar fashion.

The characteristic equation of this system is given by

$$1+zG_sH(z,h/2)-G_sH(z,1-h/2)=0 \quad (84)$$

Following the z-transform techniques discussed in the literature, the investigation of the system stability, the evaluation of the output response, and the design of the system compensation can be readily carried out.

This discussion further points out that the determination of the pulse-transfer functions $G_1(z)$ and $G_1H(z)$ of this paper is not as complicated as the authors might think. In fact, both $G_1(z)$ and $G_1H(z)$ can be readily determined from the table of modified z-transforms[3] if use is made of the relationships given in equations 81 and 82. For further discussions of the approximate analysis of this problem, the reader is referred to chapter XI of reference 4.

References

1. See reference 7 of the paper.
2. Commutated Networks. *Report 58-12*. The Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., 1957, pp. 57-86.
3. Analysis of Sampled-Data Feedback Control Systems with Finite Sampling Duration, Julius Tou. *AIEE CP no. 58-801* (available on request).
4. Digital and Sampled-Data Control Systems (book), Julius Tou. McGraw-Hill Book Company, Inc., New York, N. Y. (to be published in the spring of 1959).
5. The Pulse Transfer Function and Its Application to Sampling Servo Systems, R. H. Barker. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. IV, 1952, pp. 302-17.

Gordon J. Murphy and **Harold B. Kennedy**: We are grateful to Professor Tou for calling our attention to other work on the analysis of sampled-data systems with appreciable pulse width. However, we object to his referring to the other accounts as "previous papers." Professor Tou's reference 3 had not been presented at the time this paper was submitted, and has not yet been published. Furthermore, even the scheduled publication date for his reference 4 is yet several months in the future.

In Fig. 13, Professor Tou merely illustrates qualitatively what is stated mathematically in equation 67 of the paper under discussion, and his Fig. 14 is identical to Fig. 9(A) of the paper except for the triviality of combining the two transcendental transfer functions immediately to the right of the sampler in Fig. 9(A). Finally, the similarity of the approach suggested by the discusser to that presented in the paper is completely established in the following two steps. The substitution of equation 69 into 70 and the application of the real translation theorem yield

$$D(z)=\frac{Z\left\{e^{\frac{hs}{2}}(1-e^{-hs})G_s(s)\right\}Z\left\{e^{\frac{hs}{2}}R(s)\right\}}{1+Z\left\{e^{\frac{hs}{2}}(1-e^{-hs})G_s(s)H(s)\right\}} \quad (85)$$

where

$$G_s(s)\triangleq\frac{G(s)}{s} \quad (86)$$

as in equation 80. The application of equation 72 then gives

$$C(z)=\frac{Z\left\{(1-e^{-hs})G_s(s)\right\}Z\left\{e^{\frac{hs}{2}}R(s)\right\}}{1+Z\left\{e^{\frac{hs}{2}}(1-e^{-hs})G_s(s)H(s)\right\}} \quad (87)$$

which is identical to equation 80.

It is well known[1] that $Z[e^{\pm Ms}G(s)]$ can be expressed in terms of modified z-transforms, as in equations 81 and 82, but we chose not to use this approach. In our opinion, it is equivalent, but not superior, to the method proposed in the paper under discussion.

It is interesting to note here that obviously Professor Tou did not use equation 82 in his published study[4] of open-loop sampled-data systems with finite sampling duration. Had he done so, he would not have obtained the erroneous final results presented in equations 72 and 74 of that paper.

A more extensive treatment of the method under discussion was presented[2] at the 1958 National Electronics Conference. The interested reader is referred to that paper for an illustrative example on the use of the method and for a comparison of the results thus obtained with experimental results.

References

1. The z-Transform: A Method for Analysis of Sampled-Data Systems, C. H. Knapp, E. Shapiro, R. A. Thorpe. *Research Report RC-47*, International Business Machines Corporation, New York, N. Y., Dec. 1957, p. 39.
2. An Approximation to Simplify the Analysis of Closed-Loop Sampled-Data Systems with Non-negligible Pulse Width, Gordon J. Murphy. *Proceedings*, National Electronics Conference, Chicago, Ill., 1958.
3. See reference 7 of the paper.

Instability in a Nonlinear Conditionally Stable System Subjected to a Sinusoidal Input

J. L. DOUCE
NONMEMBER AIEE

R. E. KING
NONMEMBER AIEE

Synopsis: This paper considers the response of a saturating conditionally stable system to a sinusoidal input. When the input signal is sufficiently large, the system becomes unstable and it is found that the resultant oscillation depends to a marked extent on the frequency of the input signal. An extension to the previously developed describing-function technique is presented, which gives a unified explanation for all the observed phenomena. Experimental work agrees with theoretical predictions for a particular system.

THE describing-function technique has been widely applied for assessing the stability of nonlinear control systems.[1] For systems which become continuously oscillatory when suitably excited, the method predicts accurately the amplitude and frequency of the resultant oscillations. When more than one state of continuous oscillation is predicted, theory enables the final state of the system to be evaluated by distinguishing between stable and unstable oscillations.

A class of systems of particular interest is stable for the operating value of open-loop gain, but becomes unstable if the gain is reduced sufficiently. Fig. 1 shows the open-loop frequency-response locus for a particular conditionally stable system. When such a system is handling signals sufficiently large to saturate some element, the effective loop gain will be

Paper 58-1327, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26-31, 1958. Manuscript submitted November 15, 1957; made available for printing September 10, 1958.

J. L. Douce and R. E. King are with The Queen's University of Belfast, Belfast, Northern Ireland.

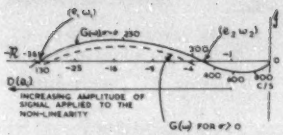


Fig. 1. Nyquist diagram of linear system G(ω) and the simple describing function D(e_i)

reduced. The particular case to be considered is where a sinusoidal signal is applied to the type of system of sufficient amplitude to drive the system into saturation. Many different modes of oscillation are observed experimentally under these conditions, and the frequency of oscillation may be synchronized to or independent of the input frequency.

It is found that extensions to the simple describing-function technique enable all the observed phenomena to be predicted.

Nomenclature

$G(\omega)$ = frequency response of the linear system
$D(e_i)$ = describing function of a nonlinear element $D(e_i) = -1/N(e_i)$
e_i = Input signal to the nonlinear element
e_o = output of the nonlinear element
$N(e_i)$ = gain of the nonlinear element
h = width of the linear regime
a, b = normalized amplitudes of signals applied to the nonlinear element

The Single-Input Describing Function

To evaluate the stability of a nonlinear closed-loop system, containing one nonlinear element, the system is first split into two parts, the linear portion, whose response is frequency-dependent but independent of amplitude, and the nonlinear portion, whose instantaneous output depends only on the input and not on its time derivatives.

It is assumed that any continuous oscillation of the closed loop will produce a sinusoidal signal at the input to the nonlinearity. This is legitimate in all practical servomechanisms, where the motor and associated inertia provide inherent low-pass filtering networks in the loop. Since the effect of distortion is neglected when evaluating the response of the nonlinearity, only the fundamental of its output need be considered, and the effective gain of the nonlinearity is defined as:

$$N_1(e_i) = \frac{\text{fundamental of output}}{\text{input amplitude}}$$

Thus the over-all loop gain is $G(j\omega)$ $N_1(e_i)$ and for continuous oscillation, $G(j\omega)\ N_1(e_i) = -1$; i.e.,

$$G(j\omega) = -\frac{1}{N_1(e_i)}$$

Hence if $G(j\omega)$ and $-1/N_1(e_i)$ are plotted on the same graph, any intersection indicates the possibility of instability of the system, and the values of ω and e_i at the point of intersection give the frequency and amplitude of the resultant oscillations.

For saturation, the describing function $D_1(e_i)$ is given by:

$$D_1(e)_i = \frac{-\pi/2}{\sin^{-1}\left(\frac{1}{m}\right) + \frac{1}{m}\left[1-\left(\frac{1}{m}\right)^2\right]^{1/2}} \quad (1)$$

where

$e_o = e_i$ for $|e_i| < h$

$e_o = h\,\frac{e_i}{|e_i|}$ for $|e_i| > h$

and

$e = mh \sin \omega t$

This function is plotted in Fig. 1.

It will be observed that there are two intersections of the frequency-dependent and amplitude-dependent loci. However, it may readily be shown that one represents a stable and the other an unstable state of oscillation.

Consider the state represented by the point $(e_2\omega_2)$, and let the amplitude increase by a small amount. This increase in amplitude will modify the response of the nonlinearity, and the operating point on the describing-function locus will move along the locus in the direction of increasing amplitude. In Fig. 1, this corresponds to a shift to the left from the point $(e_2\omega_2)$.

The modified resultant oscillation must have unity over-all loop gain. Since the operating point along the describing-function locus has been determined, the loop gain of the linear system for the resultant oscillations is known.

Writing the resultant oscillation as

$e_i = A_o e^{(\sigma + j\omega)t}$

the values of σ and ω can be determined from Fig. 1.

Linear theory shows that loci of $G(\sigma, \omega)$ for $\sigma > 0$ lie below the locus of $G(\sigma = 0, \omega)$. Thus the system has an over-all gain of unity for a signal of the form

$e_i = A_o e^{(\sigma + j\omega)t}$

where $\sigma > 0$. This corresponds to an oscillation of increasing amplitude. Hence if the amplitude of oscillation should undergo a slight increase, the response of the system increases indefinitely with time.

Conversely, any slight decrease in the amplitude of the oscillation causes the amplitude of the oscillation to decay indefinitely. This shows that the oscillation is unstable, so that the operating conditions diverge from the point $e_2\omega_2$.

Identical reasoning shows that the oscillation of amplitude e_1 and of frequency ω_1, is stable for any slight disturbance. Thus any small increase in amplitude produces a resultant oscillation of decreasing amplitude, until the initial conditions are restored.

The determination of the stability of the oscillations of a system is a problem peculiar to nonlinear systems. The most profitable approach, used throughout this paper, considers the effect of a slight disturbance on the oscillation. If the slight perturbation produces an effect which decays with time, the assumed oscillation is stable, otherwise the oscillation assumed is unstable, and will not occur in practice.

Thus the system will oscillate only at frequency ω_1 in the absence of an applied stimulus. It will be shown later how an applied sine wave of arbitrary frequency can permit the system to oscillate at frequency ω_2.

While previous theory gives information about the natural stability of the system, no prediction can be made of stability when a sinusoidal signal is applied, except for the special cases when its frequency is the frequency of the natural oscillations.

By considering the behavior of the nonlinearity when two sinusoidal signals are applied simultaneously, the complete response can be predicted.

The Dual-Input Describing Function

To assess the stability of a system in the presence of an applied sinusoid, it is necessary to consider the response of the system to a sinusoidal test signal in the presence of the primary forcing signal.

Since the superposition theorem applies

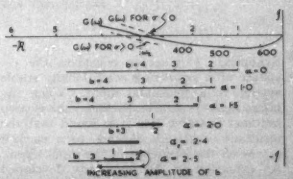


Fig. 2. Describing functions for unrelated frequencies with part of the Nyquist plot

to the linear portion of the system, the behavior of this portion is not modified by the primary input signal. However, the presence of the primary input signal can appreciably modify the response of the nonlinear element to a further test signal, and this effect has a dominating influence on the over-all response.

The primary sinusoidal signal is regarded as the response of the system as given by steady-state theory. Thus the amplitude and frequency of this signal are known, and are regarded as constant. The second signal is the test signal, injected for determining the stability of the operating conditions given by steady-state theory.

Due to the primary signal modifying the gain of the nonlinear element, the over-all closed loop can become unstable, producing stable or unstable oscillations. For stability, the over-all loop gain to the test signal must have a phase shift of less than π radians at the frequency of unity gain. More conveniently, the describing function for the test signal must not cut the open-loop frequency response locus of the linear system.

To determine the response of the nonlinearity to the test signal, it is necessary to evaluate the frequency components of the output when the input consists of two sinusoidal signals applied simultaneously. Since the test signal may be of any amplitude and frequency, the input to the nonlinearity is written

$$e_i = ah\cos(\omega t + \phi) + bh\cos n\omega t$$

The first term is regarded as the primary input, and the second term represents the test signal applied to the nonlinearity.

The output signal, e_o, will contain components of frequency ω, $n\omega$, harmonics of these frequencies, and intermodulation frequencies. The portion of the output signal of interest is that component of frequency $n\omega$, due to the test signal input.

Hence the gain of the nonlinearity is defined as

$$N_2(e_i) = \frac{\text{output amplitude of desired component}}{bh}$$

and the describing function

$$D_2(e_i) = \frac{-1}{N_2(e_i)}$$

The gain to the second signal depends on n, the relative frequency of the test signal and primary signal. Three cases are of importance. These are $n=1$, where the test signal is of the same frequency as the primary input, $n=1/3$ so that the test signal is of one third the primary input frequency, and the case where the frequency of the test signal bears no simple relationship to the primary input frequency.

These cases will be considered separately. Three separate phenomena are encountered, corresponding to the "jump" effect, subharmonic resonance, and the occurrence of oscillations unrelated to the input frequency.

Finally, a unified picture of the complete sinusoidal response is presented, showing the relative importance of each effect.

Unrelated Frequencies

When a sinusoidal signal is applied to any single-valued nonlinearity, the fundamental of the output waveform is always in phase with the input sinusoid. Similarly, when two sinusoidal signals are applied simultaneously, the components of the input frequencies in the output signal are in phase with the input sinusoidal signals, provided that the two frequencies are unrelated.

It follows that the describing function of any single-valued nonlinearity for a sinusoidal signal, in the presence of a primary applied sinusoid of unrelated frequency, will be purely real. This describing function has been evaluated for saturation using graphical techniques and a digital computer. Results are presented in Fig. 2. For clarity, the functions are shown parallel to the negative real axis, instead of along it. The signal of amplitude a is regarded as the primary signal, and the describing function for the signal of amplitude b is plotted for a range of values of b. The case $a=0$ gives the single-input describing function.

From the describing function curves, it is seen that the general effect of the presence of a primary signal on the saturating nonlinearity is to reduce the gain to a further signal.

The presence of the primary signal also has a profound effect on the manner in which the gain to the test signal varies with its amplitude. For the case $a=0$ the gain decreases continually with increasing amplitude b. However, for $a \geq 1$, the gain to the signal b is seen to increase with increasing amplitude b, for a range of values of b. In particular, for values of a in the range $2 \leq a \leq a_1$, the describing function for the test signal cuts the open-loop frequency response locus twice for $\omega=\omega_3$. One of these intersections represents a stable oscillation, and the other an unstable state. Since the amplitude a is regarded as a constant, the stability of any oscillation can be determined by considering b to be the only variable, and the technique used with the single-input describing function is applicable.

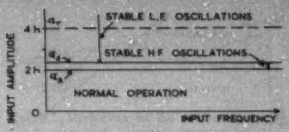


Fig. 3. Range of signals applied to the nonlinearity which excite unsynchronized oscillations

a_T is the amplitude for oscillation as predicted by the single-input describing function

Consider the intersection of the describing-function and the frequency-response locus such that $\omega=\omega_3$ and b is small, and such that any slight increase in b increases the gain of the nonlinearity. Let the amplitude b, of the oscillation, increase by a small amount. The resultant oscillation will now be of the form

$$e^{-\sigma t} e^{j\omega_3 t}$$

so that the amplitude will decrease with time until the initial conditions are restored. A small decrease of the amplitude b produces oscillations of increasing amplitude, until the previous steady state is attained.

When the intersection of the describing function and the frequency-response locus corresponds to decreasing gain with increasing b, for example the loci for $a=0$, 1.0 or 1.5, then oscillations of frequency ω_3 are unstable.

Thus stable oscillations of frequency ω_3 are possible because of the primary signal modifying the response of the nonlinearity so that, for a range of input signals, the gain to a second signal increases with increasing amplitude of this signal.

These high-frequency oscillations occur in the particular system considered when the primary signal is of amplitude in the range $2 \leq a < 2.36$. For signals of smaller amplitude, the system is stable, and the resultant response corresponds to that predicted by simple theory. When $a > 2.36$, oscillations occur of frequency ω_1 (see Fig. 1) and these oscillations persist when the primary input signal is removed.

As a comparison, the single-input describing function predicts stable operation until the signal applied to the nonlinearity is of amplitude $a=4.0$.

When the frequency of oscillation bears no simple relationship to the primary input frequency, the resultant oscillations depend only on the amplitude of the primary signal applied to the nonlinearity, and are independent of the primary input

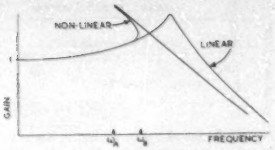


Fig. 4. Typical closed-loop frequency-response curves, showing the jump effect

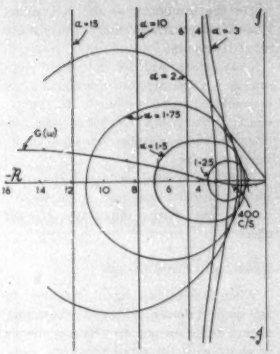


Fig. 5. Incremental describing-function loci

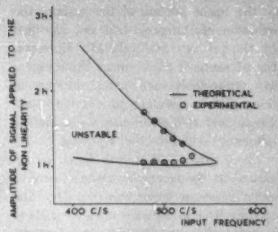


Fig. 6. Region of instability producing the jump effect

frequency. Fig. 3 shows the types of unsynchronized oscillations which can occur as a function of the amplitude and frequency of the primary input.

Input Frequencies Identical

The previous section discussed the response of a nonlinearity to a sinusoidal signal in the presence of a second signal of unrelated frequency. A further important case is when the test signal is of the same frequency as that of the primary input. In particular, the response of the nonlinearity is presented for a test signal of vanishingly small amplitude, and of the primary-input frequency.

This enables the over-all system's behavior to be found for a perturbation in the steady-state response. If this slight displacement from the steady-state operating conditions produces a response which increases indefinitely with time, the assumed steady-state conditions will be unstable, and will not be observed in a practical system.

The "jump" effect[2,3] is intimately connected with the occurrence of unstable solutions for the response as predicted by steady-state theory. Fig. 4 shows typical closed-loop frequency-response curves for a nonlinear system. Over a range of frequencies, steady-state theory predicts three possible values for the closed-loop gain. In practice, only two of these are observed, and the third solution is not a stable operating condition. Abrupt changes in closed-loop gain occur at frequencies ω_A and ω_B. Thus the frequencies at which the jump effect occurs correspond to the frequencies defining the limits of unstable predicted solutions. By evaluating the range of inputs in which unstable solutions are encountered, the presence or absence of the jump effect for various inputs can be determined.

The input to the nonlinearity is now

$$e_i = a\cos(\omega t+\phi)+b\cos\omega t$$

The component of the output of the nonlinearity due to b is now determined, and the gain found for the input component $b\cos\omega t$, as $b\to 0$. The gain is a function of a and ϕ. The corresponding describing function has been termed the incremental describing function.

Intermodulation terms produced by the nonlinear characteristic can introduce a phase shift between input and output of the desired component. As an example, consider the cubic characteristic $e_o = e_i^3$

$$e_i = a\cos(\omega t+\phi)+b\cos\omega t, \quad b\to 0$$

$$e_o = a^3\cos^3(\omega t+\phi)+b^3\cos^3\omega t+3a^2b\times\cos^2(\omega t+\phi)\cos\omega t+3ab^2\times\cos(\omega t+\phi)\cos^2(\omega t+\phi)$$

The component of the output due to the incremental signal is

$$e = 3a^2b\cos^2(\omega t+\phi)\cos\omega t+ \text{terms in } b^2 \text{ and } b^3$$

$$= \frac{3a^2b}{4}[2\cos\omega t+\cos(\omega t+2\phi)]+ \text{terms in } 3\omega$$

Thus that component of the output of input frequency due to the test signal has now undergone a phase shift in passing through the nonlinearity. It is this phase shift which may allow the resultant closed-loop system to become unstable when a sinusoidal signal is applied. For each value of a, the gain function locus describes a closed curve as ϕ is varied.

The describing function provides a more convenient means of assessing the stability of any operating condition. Fig. 5 shows the incremental describing function for saturation. Evaluation of these can be simply achieved when the relationship between fundamental of output amplitude is known for a sinusoidal input. The derivation is discussed in the Appendix. For any particular value of a, the describing-function locus for the incremental signal follows a closed curve as ϕ is varied. The gain to the incremental signal can become very small when $a\geq 3$, and it is not possible to illustrate the closed nature of these loci. For practical purposes, when $a\geq 3$, all the area to the left of the locus may be considered enclosed by the curve.

Fig. 5 also shows the open-loop frequency-response locus of the linear system, a large portion being enclosed by the incremental describing-function loci. As an example of the application of this diagram, consider the complete closed-loop system subject to a sinusoidal signal of frequency 400 cycles per second. For a range of amplitudes of the signal applied to the nonlinearity, the solution as predicted by steady-state theory will be unstable. This implies that the response to a small, incremental signal increses with time, so that the incremental describing function encloses the point on the open-loop gain locus corresponding to the input frequency. From the figure, the describing function encloses the 400-cycle-per-second gain point for a in the range $1.2<a<3$. This defines a range of a which will never be observed in a practical system, regardless of the input amplitude to the over-all system. The range of instability is shown in Fig. 6 for the system considered.

Input Frequencies Simply Related

When two sinusoidal signals of simply related frequencies are applied to a nonlinearity, the output waveform will, in general, contain components of both these frequencies. Due to intermodulation, there will usually be a phase shift between input and output components of the same frequency. This phase shift may be sufficient to make the completed closed-loop system unstable, the frequency of oscillation being simply related to the input frequency.

A case of practical importance arises when the frequency of oscillation is a

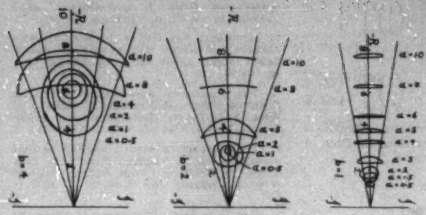


Fig. 7. Low-frequency describing function loci for N = 1/3

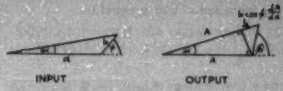


Fig. 9. Derivation of the incremental gain of a nonlinearity

submultiple of the input frequency giving low-frequency large-amplitude subharmonic oscillations.[4] For systems incorporating symmetrical nonlinearities, the only subharmonic oscillations possible are those for which the frequency of oscillation is an odd submultiple (1/3, 1/5, 1/7, etc.) of the input frequency.

The only subharmonic oscillations to be considered for the particular system are those of frequency one third the input frequency. Higher order oscillations are observed, but these occur only over a very limited range of inputs. They are explained quantitatively by an identical method.

Hence the input to the nonlinearity is considered to be

$$e_i = ah \cos(\omega t + \phi) + bh \cos \frac{1}{3}\omega t$$

It is necessary to evaluate the gain, and hence the describing function, for the component $bh \cos 1/3\omega t$. This is a function of a, b, and ϕ.

The describing function for this component always lies in a region between two lines from the $(-1, 0)$ point at 28 degrees to the negative real axis. It has been evaluated for a wide range of input parameters, using a digital computer.[5] Typical results are shown in Fig. 7.

Subharmonic oscillations occur when the input frequency to the system is of the order of three times its resonant frequency. In the particular system considered, the amplitude of the subharmonic oscillations may be sufficient to drive the system unstable such that it oscillates at its own natural frequency. In this case, when the input is removed, the free oscillations persist.

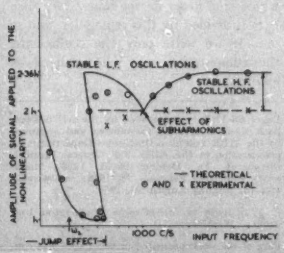


Fig. 8. Composite behavior of the system

The Complete Response

To derive the complete response of the conditionally stable system to a sinusoidal input, it is necessary to consider the region of inputs in which each separate phenomenon predominates.

From Fig. 4, it is seen that whenever the signal to the nonlinearity exceeds a certain amplitude unsynchronized oscillations will occur. For amplitudes less than a_2, these oscillations are of a high frequency, and decay if the input is removed. When the amplitude of the signal applied to the nonlinearity exceeds a_2, the oscillations are of low frequency, and persist when the input to the system is removed.

For input frequencies near to the resonant frequency of the closed-loop system, the jump effect can occur. Thus a region of error amplitudes and frequencies exists, Fig. 6, which represents an unstable system. The steady state will never correspond to a point within this region.

Finally, a high-frequency sinusoidal signal can cause subharmonic oscillations, and these oscillations may build up so that the system becomes oscillatory at its own natural frequency.

The over-all behavior is shown in Fig. 8, showing the importance of each phenomenon in affecting the stability of the conditionally stable system. Experimental work verifies the predicted effects and, as shown, gives good quantitative agreement with the theory developed.

Conclusions

It has been shown that many phenomena occur when a sinusoidal signal is applied to a conditionally stable system. As in other nonlinear systems, subharmonics and the jump effect are readily observed. Peculiar to the conditionally stable system, unsynchronized sinusoidal oscillation is possible. These unsynchronized oscillations can be either of two distinct frequencies. One of these frequencies is that predicted from the single-input describing function. The other oscillation, observed only in the presence of an applied sinusoidal signal, has a frequency corresponding to that of an unstable oscillation from previous theoretical work.

By an extension of the describing-function technique, all these different phenomena have been shown to have a common explanation. Experimental work verifies the predicted behavior of a particular system.

Appendix. Derivation of the Incremental Describing Function

From a knowledge of the normal, single-input gain function of the nonlinearity, the incremental describing function can be derived. This follows since the addition of two sinusoidal signals of the same frequency gives a further sinusoidal signal. This method is limited to single-valued nonlinearities, since it is assumed that the fundamental of the output is in phase with the input signal.

Let the primary input be of amplitude a, and its corresponding output have amplitude A. This is illustrated vectorially in Fig. 9. Consider the effect of a further small applied sinusoidal signal of amplitude b, at an angle ϕ with respect to the primary signal. The resultant input is shifted in phase by the small angle α.

The new input amplitude is $a + b \cos \phi$, and this produces an output

$$A + b \cos \phi \left[\frac{dA}{da}\right]_a = A + b \cos \phi A'$$

where A' is the slope of the curve of fundamental of output plotted against input for the input amplitude a.

This output signal is in phase with the new input signal.

Thus

$$\tan \beta = \frac{A\alpha}{A'b \cos \phi} = \frac{A}{aA'} \tan \phi$$

The incremental gain is equal to

$$|N_h| = \frac{1}{b}\left[\left(\frac{A}{a}\right)^2 b^2 \sin^2\phi + b^2(a')^2\cos^2\phi\right]^{1/2}$$

and is independent of b.

The phase shift is equal to $(\beta-\phi)$ and is given by

$$\tan(\beta-\phi) = \frac{\left(\frac{A}{aA'}-1\right)\tan\phi}{1+\frac{A}{aA'}\tan^2\phi}$$

For symmetrical saturation limiting at $\pm k$,

$$A = \frac{2k}{\pi}\left\{a\sin^{-1}\left(\frac{1}{a}\right)+\left[1-\left(\frac{1}{a}\right)^2\right]^{1/2}\right\}$$

By differentiation,

$$A' = \frac{2k}{\pi}\left\{\sin^{-1}\left(\frac{1}{a}\right)-\frac{1}{a}\left[1-\left(\frac{1}{a}\right)^2\right]^{1/2}\right\}$$

Substituting in the foregoing expressions gives the incremental describing function of Fig. 5, as a function of a and ϕ.

References

1. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms. **Ralph J. Kochenburger.** *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 270–84.

2. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements. **E. Calvin Johnson.** *Ibid.*, vol. 71, pt. II, July 1952, pp. 169–81.

3. The Frequency Response of a Certain Class of Nonlinear Feedback Systems. **J. C. West, J. L. Douce.** *British Journal of Applied Physics*, London, England, vol. 5, 1954, p. 204.

4. The Generation and Extinction of Subharmonics. **C. A. Ludeke.** *Proceedings*, Symposium on Nonlinear Circuit Analysis, Institute of Radio Engineers, New York, N. Y., 1953, p. 215.

5. The Dual Input Describing Function and Its Use in the Analysis of Nonlinear Feedback Systems. **J. C. West, J. L. Douce, R. K. Livesley.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 103, pt. B, 1956, p. 463.

Two Theorems on the Number of Real Roots of the Characteristic Equation of Any Stable Linear Physical System

J. F. KOENIG
ASSOCIATE MEMBER AIEE

THE TIME RESPONSE to a disturbance in any linear system depends upon the initial conditions, the disturbance, and the roots of the characteristic equation. The response will be bounded (stable) if, and only if, all roots have negative real parts. The response is the sum of a particular integral (PI) and a complementary function (CF). The PI is often called the steady-state part of solution and the CF the transient part of the solution.

The CF is usually of the following form

$$C_1 e^{r_1 t} + C_2 e^{r_2 t} + \ldots + C_n e^{r_n t}$$

in which the C_i are the arbitrary constants, the r_i are distinct roots of the characteristic equation of degree n, and t is the time. For each root r_i of multiplicity k, the corresponding terms in the CF are

$$(C_1 + C_2 t + \ldots + C_k t^{k-1}) e^{r_i t}$$

A stable linear system whose characteristic equation has only negative real roots will have a CF containing no oscillatory terms.

A common problem in the design of linear systems, including feedback control systems, is the solution for the roots of the characteristic equation. When solving for the roots it is often desirable to determine the number of real roots and then solve for them before solving for the complex roots. Sturm, Frank, Schmutz, Michailov, Meerov, Bloch, Kac, and Edelmann[1–8] have presented real root theorems. In the present paper a simple method is presented to determine the number of real roots of the characteristic equation of any stable linear physical system.

The present paper is related to papers[9–11] which have as their subject the location of roots within regions of the complex plane.

Distinct Real Roots

The following theorem gives the number of real roots of the characteristic equation:

Theorem 1. Let the left side of the characteristic equation be set up as follows:

$$F(s) = \sum_{k=0}^{n} a_k s^k \qquad (1)$$

a polynomial with positive real coefficients and $a_n = 1$.

With $j = \sqrt{-1}$, form the complex polynomials

$$f_0(s) = \sum_{k=0}^{n} a_k j^{n-k} s^k \qquad (2)$$

$$f_1(s) = \sum_{k=0}^{n-1} (n-k) a_k j^{n-k-1} s^k \qquad (3)$$

and the continued fraction,

$$\frac{f_1(s)}{f_0(s)} = \cfrac{1}{C_1 s + d_1 + \cfrac{1}{C_2 s + d_2 + \cdots \cfrac{1}{C_n s + d_n}}} \qquad (4)$$

in which the C_k are real (and ± 0) and the d_k pure imaginary or zero. If r of the C_k are negative then $F(s)$ has $n-2r$ negative real zeros and r pairs of complex conjugate zeros.

The proof of theorem 1 is given in a later section. To illustrate the use of theorem 1, let us apply it to the following characteristic equation:

$$F(s) = s^5 + 8s^4 + 25s^3 + 40s^2 + 34s + 12 = 0 \quad (5)$$

Substitute the coefficients of equation 5 into polynomials of equations 2 and 3 and obtain:

$$f_0(s) = s^5 + j8s^4 - 25s^3 - j40s^2 + 34s + j12 \quad (6)$$

$$f_1(s) = 8s^4 + j50s^3 - 120s^2 - j136s + 60 \quad (7)$$

Obtain the coefficients of the continued fraction in equation 4 by dividing $f_0(s)$ by $f_1(s)$ until a remainder R_1 is obtained which is of lower degree than $f_1(s)$; then $f_1(s)$ is divided by this remainder, and so on. If we write only the coefficients, the computation may be arranged as follows:

Paper 58-1267, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted February 24, 1958; made available for printing August 29, 1958.

J. F. Koenig is with George Washington University, Washington, D. C.

Part of the work presented in this paper was conducted at the United States National Bureau of Standards, Washington, D. C.

$$
\begin{array}{l}
0.125s+j0.219 = C_1s+d_1 \\
f_1(s)/1+j8.000-25.000-j40.000+34.000+j12.000 = f_0(s) \\
\quad 1+j6.250-15.000-j17.000+\ 7.500 \\
\quad +j1.750-10.000-j23.000+26.500+j12.000 \\
\quad +j1.750-10.950-j26.280+29.784+j13.140 \\
\quad\quad +\ 0.950+j\ 3.280-\ 3.284-j\ 1.140 = R_1
\end{array}
$$

$$
\begin{array}{l}
8.421s+j23.557 = C_2s+d_2 \\
R_1/8+j50.000-120.000-j136.000+60 = f_1(s) \\
\quad 8+j27.621-\ 27.655-j\ \ 9.600 \\
\quad\quad j22.379-\ 92.345-j126.400+60.000 \\
\quad\quad j22.379-\ 77.267-i\ 77.361+26.855 \\
\quad\quad\quad -\ 15.078-i\ 49.039+33.145 = R_2
\end{array}
$$

$$
\begin{array}{l}
-0.0630s-j0.0127 = C_3s+d_3 \\
R_2/0.950+j3.280-3.284-j1.140 = R_1 \\
\quad 0.950+j3.089-2.088 \\
\quad\quad +j0.191-1.196-j1.140 \\
\quad\quad +j0.191-0.623-j0.421 \\
\quad\quad\quad -0.573-j0.719 = R_3
\end{array}
$$

$$
\begin{array}{l}
26.314+j52.564 = C_4s+d_4 \\
R_3/-15.078-j49.039+33.145 = R_2 \\
\quad -15.078-j18.920 \\
\quad\quad -j30.119+33.145 \\
\quad\quad -j30.119+37.794 \\
\quad\quad\quad -\ 4.649 = R_4
\end{array}
$$

$$
\begin{array}{l}
0.123-j0.155 = C_5s+d_5 \\
R_4/-0.573-j0.719 = R_3 \\
\quad -0.573-j0.719 \\
\quad\quad 0 = R_5
\end{array}
\qquad (8)
$$

Four of the C_k are positive and one is negative. Therefore, by theorem 1, there are $n-2r=5-2=3$ real roots. This is correct, because the characteristic equation 5 was formed from the roots -1, -2, -3 and $-1\pm j1$.

The following theorem determines whether all roots are real, distinct, and negative. The proof is given in a later section.

Theorem 2. All zeros of the real polynomial

$$F(z)=\sum_{k=0}^{n} a_k z^k \qquad (1)$$

are real, negative, and distinct if, and only if, all coefficient C_k are positive and the d_k are pure imaginary or zero in the continued fraction expansion

$$\frac{f_1(z)}{f_0(z)} = \cfrac{1}{C_1s+d_1+\cfrac{1}{C_2s+d_2+\ldots\cfrac{1}{C_ns+d_n}}} \qquad (4)$$

in which

$$f_0(s)=\sum_{k=0}^{n} a_k j^{n-k} z^k \qquad (2)$$

$$f_1(s)=\sum_{k=0}^{n-1} (n-k)a_k j^{n-k-1} z^k \qquad (3)$$

Note that the polynomials 1, 2, and 3 and continued fraction 4 of theorem 2 are the same as in theorem 1. Let us apply theorem 2 to the following characteristic equation:

$$F(s)=s^5+15s^4+85s^3+225s^2+274s+120=0 \qquad (9)$$

Form

$$f_0(s)=s^5+j15s^4-85s^3-j225s^2+274s+j120 \qquad (10)$$

and

$$f_1(s)=15s^4+j170s^3-675s^2-j1096s+600 \qquad (11)$$

Obtain the coefficients of continued fraction 4 as follows:

$$
\begin{array}{l}
0.067s+j0.244 = C_1s+d_1 \\
f_1(s)/1+j15.000-85.000-j225.000+274.000+j120.000 = f_0(s) \\
\quad 1+j11.333-45.000-j\ 73.067+\ 40.000 \\
\quad\quad +j\ 3.667-40.000-j151.933+234.000+j120.000 \\
\quad\quad +j\ 3.667-41.480-j164.700+267.424+j146.400 \\
\quad\quad\quad +\ 1.480+j\ 12.767-\ 33.424-j\ 26.400 = R_1
\end{array}
$$

$$
\begin{array}{l}
10.135s+j27.436 = C_2s+d_2 \\
R_1/15+j170.000-675.000-j1096.000+600.000 = f_1(s) \\
\quad 15+j129.394-338.752-j\ 267.564 \\
\quad\quad +j\ 40.606-336.248-j\ 828.436+600.000 \\
\quad\quad +j\ 40.606-350.275-j\ 917.021+724.310 \\
\quad\quad\quad +\ 14.027+j\ \ 88.585-124.310 = R_2
\end{array}
$$

$$
\begin{array}{l}
0.106s+j0.241 = C_3s+d_3 \\
R_2/1.480+j12.767-33.424-j26.400 = R_1 \\
\quad 1.480+j\ 9.390-13.177 \\
\quad\quad j\ 3.377-20.247-j26.400 \\
\quad\quad j\ 3.377-21.349-j29.959 \\
\quad\quad\quad +\ 1.102+j\ 3.559 = R_3
\end{array}
$$

$$
\begin{array}{l}
12.729s+j39.276 = C_4s+d_4 \\
R_3/14.027+j88.585-124.310 = R_2 \\
\quad 14.027+j45.303 \\
\quad\quad +j43.282-124.310 \\
\quad\quad +j43.282-139.783 \\
\quad\quad\quad +\ 15.473 = R_4
\end{array}
$$

$$
\begin{array}{l}
0.071+j0.230 = C_5s+d_5 \\
R_4/1.102+j3.559 = R_3 \\
\quad 1.102+j3.559 \\
\quad\quad 0 = R_5
\end{array}
\qquad (12)
$$

Since all C_k are positive all roots are real, negative, and distinct. This result is correct because the equation has the roots -1, -2, -3, -4, and -5.

Multiple Real Roots

In case there are multiple roots (real or complex) they can be solved for in the following manner.

Given:

$$F(z)=\sum_{k=0}^{n} a_k z^k \qquad (1)$$

$$
\begin{aligned}
f_0(z)&=\sum_{k=0}^{n} a_k j^{n-k} z^k = j^n \sum_{k=0}^{n} a_k (-jz)^k \\
&= j^n F(-jz) \qquad (2A)
\end{aligned}
$$

$$
\begin{aligned}
f_1(z)&=\sum_{k=0}^{n} a_k j^{n-k+1}(k-n) z^k \\
&= -\sum_{k=0}^{n} k a_k j^{n-k+1} z^k - n\sum_{k=0}^{n} a_k j^{n-k+1} z^k \\
&= jz\sum_{k=0}^{n} k a_k j^{n-k} z^{k-1} - nj\sum_{k=0}^{n} a_k j^{n-k} z^k \\
&= jzf_0'(z) - njf_0(z) \qquad (3A)
\end{aligned}
$$

If the real polynomial 1 has a zero of multiplicity two, each of value z_1, then polynomial 2(A) will have a double zero, each of value jz_1. Also note that polynominal 3(A) is expressed in terms of polynomial 2(A) and its derivative. When polynomial 2(A) has a zero of multiplicity two its derivative will have the same zero (but of multiplicity one). It is concluded that if equation 1 has a double zero then equations 2(A) and 3(A) will have a common factor which can be found by application of the Euclidean algorithm which determines the greatest common divisor of any two polynomials. To illustrate, suppose that the characteristic equation is

$$F(z)=s^6+14s^4+75s^3+190s^2+224s+96=0 \qquad (13)$$

and we wish to determine the number of real roots, including multiple ones. Let

us attempt to obtain the coefficients of continued fraction expansion 4. The procedure of the Euclidean algorithm is identical with that set forth in the calculation of theorems 1 and 2. We form:

$$f_0(z)=z^6+j14z^4-75z^3-j190z^2+224z+j96 \tag{14}$$

and

$$f_1(z)=14z^4+j150z^3-570z^2-j896z+480 \tag{15}$$

and

$$\begin{array}{r}
+z+j3.286=C_1z+d_1\\
f_1(z)/1+j14.000-75.000-j190.000+224.000+j96=f_0(z)\\
\underline{1+j10.714-40.714-j\ 64.000+\ 34.286}\\
+j\ 3.286-34.286-j126.000+189.714+j\ 96.000\\
\underline{+j\ 3.286-35.206-j133.786+210.304+j112.664}\\
+\ 0.920+j\ \ 7.786-\ 20.590-j\ 16.664=R_1
\end{array}$$

$$\begin{array}{r}
+1.087z+j2.447=C_2z+d_2\\
R_1/1+j10.714-40.714-j64.000+34.286=f_1(z)\\
\underline{1+j\ 8.463-22.381-j18.114}\\
+j\ 2.251-18.333-j45.886+34.286\\
\underline{+j\ 2.251-19.052-j50.384+40.777}\\
+\ 0.719+j\ 4.498-\ 6.491=R_2
\end{array}$$

$$\begin{array}{r}
+1.280z+j2.822=C_3z+d_3\\
R_2/0.920+j7.786-20.590-j16.664=R_1\\
\underline{0.920+j5.757-\ 8.308}\\
+j2.029-12.282-j16.664\\
\underline{+j2.029-12.693-j18.318}\\
+\ 0.411+j\ 1.654=R_3
\end{array}$$

$$\begin{array}{r}
+1.749z+j3.905=C_4z+d_4\\
R_3/0.719+j4.498-6.491=R_2\\
\underline{0.719+j2.893}\\
+j1.605-6.491\\
\underline{+j1.605-6.459}\\
-0.032=R_4
\end{array} \tag{16}$$

Here we find that R_4 is negligibly small, so that C_4 and d_4 do not appear. If R_4 were exactly zero then the Euclidian algorithm states that the greatest common divisor of $f_1(z)$ and $f_0(z)$ would be R_3 and the root we seek would be found by equating R_3 to zero. We therefore equate R_3 to zero,

$$R_3=0.411z+j1.654=0 \tag{17}$$

and obtain $z\cong -j\ 4.024$ as the common zero of the polynomials in equations 14 and 15. The multiple root of equation 13 is therefore approximately -4.024. Equation 13 was formed from the roots $-1, -2, -3, -4, -4$.

The signs of the coefficients of calculation 16 also have significance because the factor corresponding to one zero of the double zero cancels out of equations 14 and 15, and theorem 2 can be applied to the equation 16. Since the four C_k are positive the four remaining roots $(-1, -2, -3, -4)$ are distinct and are negative real roots. If, in equation 16, some of the C_k were negative then theorem 1 would apply to determine the number of negative real roots of the characteristic equation. If the characteristic equation has a root of multiplicity greater than two, a similar procedure is used.

Proof of Theorem 1

The following theorem is by Frank.[2]

Theorem 3. The zeros of the complex polynomial

$$P(z)=z^n+(A_1+jB_1)z^{n-1}+\ldots+(A_n+jB_n), \quad (A_i \text{ and } B_i \text{ real}) \tag{18}$$

all have negative real parts if and only if all coefficients C_k are positive and the k_i are pure imaginary or zero in the continued fraction expansion

$$\frac{Q(z)}{T(z)}=\cfrac{1}{C_1z+k_1+\cfrac{1}{C_2z+k_2+\ldots\cfrac{1}{C_nz+k_n}}} \tag{19}$$

where

$$Q(z)=A_1z^{n-1}+jB_1z^{n-2}+A_2z^{n-3}+jB_2z^{n-4}+\ldots \tag{20}$$

and

$$T(z)=z^n+jB_1z^{n-1}+A_2z^{n-2}+jB_2z^{n-3}+A_3z^{n-4}+\ldots \tag{21}$$

Consider a real polynomial

$$p^n+a_1p^{n-1}+a_2p^{n-2}+\ldots+a_n=0,\ a_i>0 \tag{22}$$

Substitute $p=ze^{-jg}$ and obtain

$$z^n+a_1e^{jg}z^{n-1}+a_2e^{j2g}z^{n-2}+\ldots+a_ne^{jng}=0 \tag{23}$$

The transformation $p=ze^{-jg}$ relates roots of the p-equation to roots of the z-equation, i.e., relates the p-plane to the z-plane. If the p-axes and z-axes are considered to be coincident, then this substitution rotates the z-roots counterclockwise by an angle g with respect to the p-roots. Suppose the z-roots are then made coincident with the p-roots by rotating the z-plane clockwise through g degrees with respect to the p-axes.

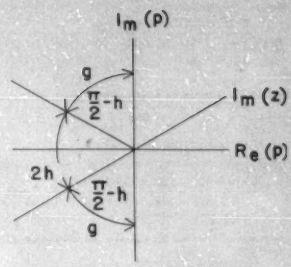


Fig. 1. Superposed complex planes

Subsequent discussion relates to this latter configuration, shown in Fig. 1. Suppose theorem 3 is applied successfully (the C_k real and positive and the k_k pure imaginary or zero) to equation 23. Then all roots of equation 23 have negative real parts and all roots of equation 22 are located within the sector $2h$, Fig. 1, since p-roots and z-roots now coincide, and complex p-roots always occur in conjugate pairs.

Equation 23 can be rewritten as:

$$\begin{array}{l}
z^n+a_1(\cos g+j\sin g)z^{n-1}+a_2(\cos 2g+\\
\qquad j\sin 2g)z^{n-2}+\ldots+a_n(\cos ng+\\
\qquad j\sin ng)=0
\end{array} \tag{24}$$

Substitute: $g=(\pi/2-h)$ into equation 24 and apply the trigonometric identities:

$$\begin{array}{l}
a_v\cos vg=a_v\cos v(\pi/2-h)\equiv\\
\quad a_v(\cos v\pi/2\cos vh+\sin v\pi/2\sin vh)
\end{array} \tag{25}$$

and

$$\begin{array}{l}
a_v\sin vg=a_v\sin v(\pi/2-h)\equiv\\
\quad a_v(\sin v\pi/2\cos vh-\cos v\pi/2\sin vh)
\end{array} \tag{26}$$

We obtain (for h approaching zero):

$$\begin{array}{l}
z^n+\Sigma a_x(xh+j1)z^{n-x}+\Sigma a_y(-1+jyh)z^{n-y}\\
x=1,\ 5,\ 9\ldots y=2,\ 6,\ 10\ldots\\
+\Sigma a_r(-rh-j1)z^{n-r}+\Sigma a_u(1-juh)z^{n-u}\\
r=3,\ 7,\ 11\ldots u=4,\ 8,\ 12\ldots
\end{array} \tag{27}$$

in which only those values of x, y, r, and u are used which correspond to nonnegative values of $n-x$, $n-y$, $n-r$, and $n-u$.

From equation 27 we form the polynomials:

$$\begin{array}{l}
T(z)=z^n+\Sigma a_x(+j1)z^{n-x}+\Sigma a_y(-1)z^{n-y}\\
x=1,\ 5,\ 9\ldots y=2,\ 6,\ 10\ldots\\
+\Sigma a_r(-j1)z^{n-r}+\Sigma a_u(1)z^{n-u}\\
r=3,\ 7,\ 11\ldots u=4,\ 8,\ 12\ldots
\end{array} \tag{28}$$

and

$$\begin{array}{l}
Q(z)=\Sigma a_x(xh)z^{n-x}+\Sigma a_y(jyh)z^{n-y}\\
x=1,\ 5,\ 9\ldots y=2,\ 6,\ 10\ldots\\
+\Sigma a_r(-rh)z^{n-r}+\Sigma a_u(-juh)z^{n-u}\\
r=3,\ 7,\ 11\ldots u=4,\ 8,\ 12\ldots
\end{array} \tag{29}$$

Polynomials 28 and 29 correspond to h approaching zero (sector approaching zero). Under this condition only negative real roots of equation 22 are located within the sector. If continued fraction 19 is formed (using polynomials 28 and 29) and theorem 3 applied, then the signs of the C_p's will indicate the number of negative real roots. h in equation 29 is a very small positive number. It can be factored out of equation 29 and will not affect the signs of the C_p's in continued equation 19. The proof of theorem 1 is now complete.

Proof of Theorem 2

The proof of theorem 2 is very similar to the proof of theorem 1. In the proof of theorem 2, the following theorem by Frank[2] is used in place of theorem 3.

Theorem 4. The complex polynomial in equation 18 has r zeros with positive real parts and $(n-r)$ with negative real parts if, in the expansion of equation 19, r of the coefficients C_p are negative and the remaining $(n-r)$ are positive.

Conclusion

A simple method is presented for the determination of the number of real roots of the characteristic equation of any stable linear physical system. The stability of the linear physical system is first determined by any standard stability test. Of course, the method presented herein does give the stability when all roots are real and negative.

References

1. Elementary Theory of Equations (book), L. E. Dickson. John Wiley & Sons, Inc., New York, N. Y., 1914, pp. 95–103.

2. On the Zeros of Polynomials With Complex Coefficients, E. Frank. *Bulletin*, American Mathematical Society, New York, N. Y., vol. 52, 1946, pp. 144–57.

3. Koeffizientenbedingungen zur Kontrolle des Dampfungsgrades bei Ausgleichsvorgangen (Verallgemeinerte Hurwitzbedingungen), O. Schmutz. *Ingenieur-Archiv*, Berlin, Germany, vol. 21, 1953, pp. 33–41.

4. Aperiodic Criterion for Automatic Regulation Systems, A. V. Michailov, *Avtomatika i Telemekhanika*, Moscow, USSR, no. 1, 1941.

5. Criterion for Aperiodicity of a Regulator, M. V. Meerov. *Izvestiya Akademii Nauk SSSR*, "Technical Sciences Section," Moscow, USSR, 1945, pp. 1169–78.

6. On Aperiodic Stability of Linear Systems, Z. Sh. Bloch. *Avtomatika i Telemekhanika*, vol. 10, no. 1, 1949, pp. 3–7.

7. On the Question of a Criterion for Aperiodic Stability (in Russian), A. M. Kac. *Prikladnaya Matematika i Mekhanika*, Moscow, USSR, vol. 15, 1951, p. 120.

8. A Criterion for the Aperiodic or the Purely Oscillatory Stability and the Relationship to Reactance Theorems, H. Edelmann. *Regelungstechnik*, Munich, Germany, no. 11, 1954, pp. 264–67.

9. A Relative Damping Criterion for Linear Systems, J. F. Koenig. *AIEE Transactions*, vol. 72, pt. II, Nov. 1953, pp. 291–95.

10. On the Zeros of Polynomials and the Degree of Stability of Linear Systems, J. F. Koenig. *Journal of Applied Physics*, New York, N. Y., vol. 24, no. 4, 1953, pp. 476–82.

11. A New Application of the Hurwitz-Routh Stability Criteria, Theron Usher, Jr. *AIEE Transactions*, vol. 76, pt. I, Nov. 1957, pp. 530–33.

Positive-Feedback Phase-Space Trajectories and Application to Servo Systems

Z. H. MEIKSIN
ASSOCIATE MEMBER AIEE

Synopsis: The phase-space trajectories for second- and third-order positive-feedback servo systems are investigated. These trajectories, together with the negative-feedback phase-space trajectories, are used to design fast-response servo systems, which switch back and forth between positive and negative feedback modes. The approach used for the third-order systems may be applied to higher order systems. Experimental results of simulated systems are presented.

THE EARLIER WORK done to shorten the response time of servo systems, by means of nonlinear and switching technique, dealt with the variation of the damping coefficient of the system.[1,2] McDonald[1] and Hopkin[3] introduced the concept of a system which is driven by a constant maximum force which reverses direction instantaneously. They applied this method to second-order systems. The system requires to switch the direction of the force by a condition which is expressed by a nonlinear relation between the error of the system and its first derivative. The same concept was extended to third-order systems by Bogner and Kazda.[4] As pointed out by McDonald,[1] this system is inherently erratic near the rest position. The slightest error will cause the application of maximum correcting force. To avoid oscillations it was suggested to switch the system into a nonlinear damped mode in the neighborhood of the rest position. This complicates the system physically.

Flügge-Lotz and Wunch[5] investigated switching-type servo systems with variable derivative and proportional feedback, under conditions which permitted the assumption that, at all times, the output was nearly equal to the input. The criteria was to obtain response with minimum time average error. The optimum values of the feedback coefficients were obtained by simulation on an analog computer.[6]

It is the purpose of this paper to investigate second-order systems (Fig. 1) and third-order systems (Fig. 2), which operate with direct positive feedback or with direct negative feedback according to predetermined conditions. The feedback constants $+1$ and -1 will be considered; these should not be regarded as the only possibilities. It will thus be attempted to investigate the effects of an instantaneous reversal of sign of the error coefficient in the differential equation which describes the control system. The effect of the coefficient of the error-rate term (associated with damping or derivative control) in conjunction with positive and negative feedback will be indicated.

This type of system proves applicable in particular to positioning-type switching servos. Special emphasis is given to a criterion which requires minimum response time with no overshoot as well as simple switching conditions.

Since with positive feedback the actuating signal is given by the sum of the

Paper 58-1301, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 30, 1958; made available for printing September 8, 1958.

Z. H. Meiksin is with the University of Pittsburgh, Pa., and Westinghouse Electric Corporation, East Pittsburgh, Pa.

This paper is a partial abstract of a thesis submitted by the author to the Graduate School of the University of Pittsburgh in partial fulfillment of the requirements for the Doctor of Philosophy degree.

The advice and assistance of Dr. D. J. Ford in the work leading to the preparation of this paper are gratefully acknowledged.

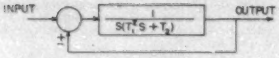


Fig. 1. Second-order servo system

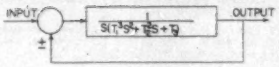


Fig. 2. Third-order servo system

input and output values, the actuating signal is relatively large and the system may operate for a portion of the transient period under maximum correcting force, which is obtained when the system saturates. With a small error, on the other hand, the actuating signal is relatively small and oscillations in the neighborhood of the rest point are inherently avoided. To decrease response time, high damping is introduced when the system switches into negative feedback. This also provides a direct simple switching condition which is expressed by a linear relationship between the error of the system and the error rate.

Second-Order Systems

The phase-plane trajectories of the system with negative feedback and the system with positive feedback are plotted. Superimposition of both sets of trajectories on one set of co-ordinates leads to the establishment of the required switching conditions to meet a prescribed criteria of system response.

Position Step Input—Negative Feedback

Error Phase-Plane Representation

The second order system may be represented by its differential equation.

$$\ddot{\theta}_e+2\zeta\omega_n\dot{\theta}_e+\omega_n^2\theta_e=0 \tag{1}$$

With application of negative feedback to the system, the sign associated with the error term is positive. The phase-plane representation of the response of the system has been widely applied to the study of servomechanisms and other feedback systems. The method is discussed in standard textbooks.[7,8] Three representative phase-plane plots of stable negative-feedback systems are shown in Fig. 3 (the error-rate ordinate has been normalized with respect to the undamped natural frequency of the system).

Eigenvectors in the Error Phase Plane

For the sake of generalization it is convenient to normalize the error rate with respect to the undamped natural frequency of the system.

Denoting

$$\dot{\theta}_{e1}=\frac{\dot{\theta}_e}{\omega_n} \text{ and } \frac{d\dot{\theta}_{e1}}{d\theta_e}=\alpha$$

equation 1 may be rewritten in the form

$$\alpha+2\zeta+\frac{\theta_e}{\dot{\theta}_{e1}}=0 \tag{2}$$

α is the slope of the error trajectories. The isocline for any given α is obtained from equation 2.

$$\frac{\dot{\theta}_{e1}}{\theta_e}=-\frac{1}{2\zeta+\alpha} \tag{3}$$

For a straight-line trajectory, $\dot{\theta}_{e1}/\theta_e=\alpha_e$. Substituting this condition into equation 3 results in an equation for the slopes of these trajectories (which are the eigenvalues or roots of the characteristic equation).

$$\alpha_e^2+2\zeta\alpha_e+1=0 \tag{4}$$

α_e is real for $\zeta\geqq 1$.

Advantage of this may be taken in determining the switching rules.

Position Step Input—Positive Feedback

Error Phase-Plane Representation

The differential equation for a positive-feedback system is, in terms of the actuating signal,

$$\ddot{\theta}_e+2\zeta\omega_n\dot{\theta}_e-\omega_n^2\theta_e=0 \tag{5}$$

The sign associated with the signal term in this equation is always negative. To compare the positive-feedback trajectories with the negative-feedback trajectories, both sets have to be plotted on one set of co-ordinates.

Applying the relation

$$\theta_e=2\theta_i-\theta_s \tag{6}$$

and the derivatives thereof to equation 5 yields the system equation in terms of the error

$$\ddot{\theta}_e+2\zeta\omega_n\dot{\theta}_e-\omega_n^2(\theta_e-2\theta_i)=0 \tag{7}$$

The error trajectories may thus be obtained from the actuating signal trajectories by translating the origin of the co-ordinates to point $(-2\theta_i, 0)$. The generality of the actuating signal trajectories, which are independent of the signal magnitude, is lost. A family of trajectories representing equation 7 is shown in Fig. 4. These curves are obtained easily by the method of isoclines, similarly to the conventional method used for negative-feedback systems. There are two straight-line trajectories (eigenvectors). Their slopes are given by

$$\alpha_e^2+2\zeta\alpha_e-1=0 \tag{8}$$

The other trajectories approach these asymptotically which facilitates the plotting of the trajectories.

As ζ is decreased, both eigenvectors rotate counterclockwise. For $\zeta=0$, they make respectively $+45$ and -45 degrees with the θ_e axis. As ζ approaches $(-\infty)$, the angle of the eigenvector with the positive slope approaches 90 degrees (the eigenvalue approaches ∞) and the angle of the other eigenvector approaches 0 degrees (the eigenvalue approaches 0).

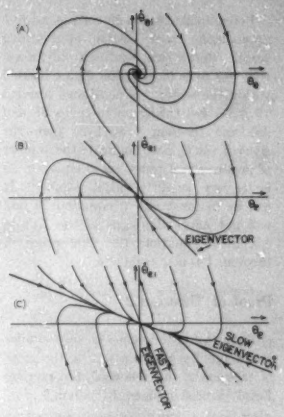


Fig. 3. Phase-plane trajectories for second-order negative-feedback systems

A—Damping factor $\zeta<1$
B—Damping factor $\zeta=1$
C—Damping factor $\zeta>1$

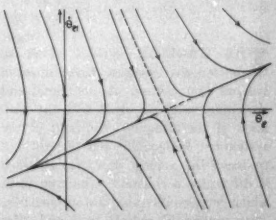


Fig. 4. Phase-plane trajectories for second-order positive-feedback system; error rate axis is normalized with respect to ω_n

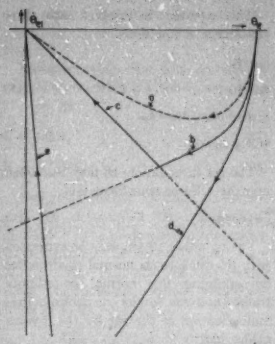


Fig. 5. Superimposition of second-order negative- and positive-feedback systems trajectories. Curves a, c, e = negative feedback. Curves b, d = positive feedback

As ζ approaches $(+\infty)$, on the other hand, the respective angles become 0 degrees and -90 degrees.

Application of Positive Feedback to Improve System Performance

The improvement of system performance by the use of positive feedback is demonstrated in Fig. 5, where the error trajectories for negative feedback and those for positive feedback are superimposed on one set of co-ordinates. In order to determine the effect of the sign of the feedback factor alone, the damping factor is kept unchanged throughout the transient period. The particular damping factor chosen is $\zeta = 1$.

An accepted criterion for the study of contactor servomechanisms is the determination of the response time of the system with no overshoot. To reach the steady state, $\dot{\theta}_e = \theta_e = 0$, the system has to approach this state along a negative-feedback trajectory. Trajectory a in Fig. 5 represents the transient response to a step input function of a negative-feedback system starting from rest. Trajectories b and c represent the transient response of the same system, except that it is connected for positive feedback during the earlier part of the transient period, and for negative feedback during the later part. The switching from the first mode of operation to the second mode takes place at the instant which corresponds to the intersection of the positive-feedback trajectory and the negative-feedback trajectory with slope α_s. This secures that the system still reaches its steady state without overshoot. The steady state is reached in a considerably shorter period than in the pure negative-feedback case. The reduction in response time is evident, because the time required by the system to change from one state to another is inversely related to the area enclosed by the θ_e axis, the trajectory, and the two respective normals to the θ_e axis which pass through the points that represent the two states.

The response time may be further decreased by having as little damping as practical during the positive-feedback mode of operation, and as much damping as practical during the negative-feedback mode of operation. This is deduced from the phase-plane characteristics as signified by the directions of the eigenvectors. The improvement of response time is illustrated by trajectories d and e in Fig. 5.

The switching along a straight line provides a simple physically realizable switching circuit. The switching condition is given by

$$\dot{\theta}_{e1} + \alpha_s \theta_e \lesseqgtr 0 \tag{9}$$

The same switching condition is valid for any magnitude of step input as shown in Fig. 6.

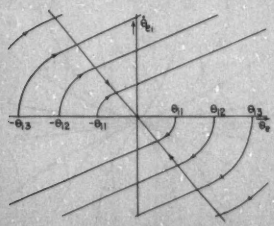


Fig. 6. Positive-feedback trajectories for several step input magnitudes, and negative-feedback eigenvector

The simple switching condition of equation 9, which was illustrated for initial conditions $\theta_{e0} = \theta_i > 0$, $\dot{\theta}_{e0} = 0$, is valid for any initial condition $0 < \theta_{e0} \leq \theta_i$ within the area bounded by the error-rate axis, the negative-feedback eigenvector, and both positive-feedback eigenvectors. This condition may be extended directly, by symmetry, to initial conditions $0 > \theta_{e0} \geq \theta_i$.

In general, the switching network becomes more complicated as the number of regions of the phase plane, in which possible initial conditions may be located, increases. For example, if $\theta_i > 0$, a switching condition must exist to secure negative-feedback operation in a region to the right of the positive-feedback eigenvector with the negative slope and below the positive-feedback eigenvector with positive slope.

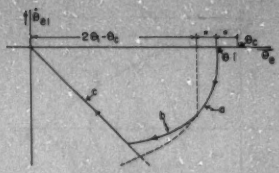


Fig. 7. Phase-plane trajectories for second-order saturating systems

Saturating Systems

In switching-type servo systems it is desired to exploit the system components to their full capacity including their saturation regions. The positive-feedback-type control described in this paper is applicable with advantage to saturating systems as illustrated in Fig. 7.

In the saturation region of the system, the actuating torque is constant, independent of the error. The torque is equal to $K\theta_c$ where θ_c is the critical signal value which causes saturation.

The equation of the system is then

$$\ddot{\theta}_e + 2\zeta\omega_n\dot{\theta}_e + \omega_n^2\theta_e = 0 \tag{10}$$

Normalizing θ_e with respect to the undamped natural frequency ω_n, dividing throughout by $\dot{\theta}_s = d\theta_s/dt$, and denoting $\dot{\theta}_{s1} = \dot{\theta}_s/\omega_n$, equation 10 becomes:

$$\frac{d\dot{\theta}_{s1}}{d\theta_s} + 2\zeta + \frac{\theta_e}{\dot{\theta}_{s1}} = 0 \tag{11}$$

from which the isoclines are obtained

$$-\dot{\theta}_{s1} = \frac{\theta_e}{2\zeta + \alpha} \tag{12}$$

where

$$\alpha = \frac{d\dot{\theta}_{s1}}{d\theta_s}$$

The isoclines are parallel to the θ_e axis.

For the positive-feedback connection, the system will saturate when

$$\theta_i + \theta_e \geq \theta_c \tag{13}$$

This corresponds to the error

$$\theta_{e0} \leq 2\theta_i - \theta_c \tag{14}$$

In Fig. 7, the system starts to respond in the positive-feedback mode, saturates when equation 14 holds, and switches into the negative-feedback mode when $\dot{\theta}_{s1} + \alpha_s\theta_e = 0$. The same system, when connected strictly for negative feedback, would not be driven into saturation.

Ramp Function Input

Negative Feedback

The error phase-plane trajectories for a ramp function input may be represented

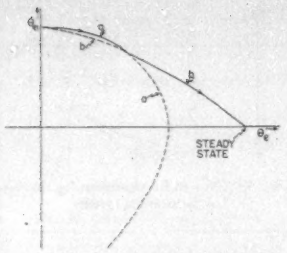


Fig. 8. Ramp-function input. Second-order phase-plane trajectories

by the error trajectories of the system with position step input function, if the origin of the co-ordinates is translated to point

$$\left(-2\frac{\zeta}{\omega_n}\dot{\theta}_i,\ 0\right) \tag{15}$$

Positive Feedback

In analogy to the negative-feedback system, the actuating signal trajectories for positive feedback with ramp function input, may be represented by the actuating signal trajectories for position step input, if the origin of the co-ordinates is translated to point

$$\left(+2\frac{\zeta}{\omega_n}\dot{\theta}_i,\ 0\right)$$

The error trajectories, θ_e, are obtained from the actuating signal trajectories, θ_i, by $\theta_e=2\theta_i-\theta_s$ and the first derivatives. Since $\dot{\theta}_i$ is constant, the $\dot{\theta}_s$ values are obtained directly by symmetry. The values of $\dot{\theta}_s$ are obtained by setting

$$\theta_i=\dot{\theta}_i t \tag{16}$$

The time t corresponding to any θ_s is obtained by conventional methods from the actuating-signal phase plane.

SWITCHING MODES FOR RAMP INPUT

An example is shown in Fig. 8. The system starts to respond from rest. Operation in the positive-feedback mode for a short interval decreases the response time. The improvement is, however, relatively small. To reach steady state, the error rate must reduce to zero and the error to a finite value. With positive feedback the error rate will approach zero faster. But at the same time, after having reached some maximum value, the error will start to decrease without ever having reached the required steady-state value. The system has, therefore, to switch into the negative-feedback mode early in the transient period, before the error starts to decrease.

To secure fastest response, the system has to choose at any state the trajectory with the higher slope, which leads towards the required steady state. This requires a switching mechanism with stored information as to the slope of which mode is larger for a finite, but large enough, number of points on the phase plane. This technique may not prove very practical.

OTHER INPUT FUNCTIONS

The switching condition expressed in equation 9 is directly applicable to a position jump staircase function input. Switching conditions for input functions, which may be approximated satisfactorily by sectional linearization, may be developed by considering them as combinations of step and ramp functions. The switching network will, however, in general be more complicated.

Third-Order System

The third-order system with three distinct real roots will be investigated for a position step input. (Distinct real roots are assumed for simplicity. A similar approach may be taken for complex and multiple roots.) The differential equation is resolved into three simultaneous equations. The system is described in the 3-dimensional phase space. To facilitate analytical treatment of the system trajectories, they are transformed to their normal phase space by means of the modal matrix,[9] and projected on two normal planes. The advantage of this change in reference co-ordinates lies in the fact that the projections of the trajectories on any of the three normal planes are independent of the third co-ordinate. Vector representation of linear differential equations in the phase space[10,11] and the normal co-ordinate transformation has been applied to the study of servo systems[4,12,13] and are assumed to be known by the reader.

It should be noted that the respective normal phase spaces of the negative-feedback and positive-feedback trajectories are not identical. The trajectories will be first obtained for the negative-feedback, then for the positive-feedback connected system.

NEGATIVE FEEDBACK

A stable system is described by the following differential equation with positive coefficients.

$$\dddot{\theta}_e+a\ddot{\theta}_e+b\dot{\theta}_e+c\theta_e=0 \tag{17}$$

This equation is resolved into three simultaneous equations.

$$\begin{aligned}\theta_{e1}&=\dot{\theta}_{e1}\\ \dot{\theta}_{e1}&=\dot{\theta}_{e2}\\ -c\theta_{e1}-b\theta_{e2}-a\theta_{e3}&=\dot{\theta}_{e3}\end{aligned} \tag{18}$$

where $\theta_{e1}=\theta_e$

The set in equation 18 may be looked upon as a linear transformation

$$\|A\|\theta_e=\dot{\theta}_e \tag{19}$$

The equation 19 set may be expressed with reference to its normal co-ordinates by applying the normal co-ordinate transformation to its variables. Premultiplication of the set by the inverse modal matrix

$$\|M\|^{-1}\|A\|\|M\|\theta_e=\|M\|^{-1}\|M\|\dot{\theta}_e \tag{20}$$

results in three independent equations referred to the normal co-ordinates, as given in equations 21 to 23.

$$\dot{X}_{e1}-\lambda_1 X_{e1}=0 \tag{21}$$

$$\dot{X}_{e2}-\lambda_2 X_{e2}=0 \tag{22}$$

$$\dot{X}_{e3}-\lambda_3 X_{e3}=0 \tag{23}$$

λ_1, λ_2, and λ_3 are the roots of the characteristic equation of the set in equation 18. They have all-negative values.

The solutions of these equations are respectively

$$X_{e1}=X_{e10}\exp\lambda_1 t \tag{24}$$

$$X_{e2}=X_{e20}\exp\lambda_2 t \tag{25}$$

$$X_{e3}=X_{e30}\exp\lambda_3 t \tag{26}$$

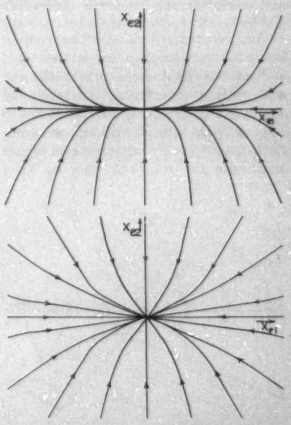


Fig. 9. Third-order negative-feedback system. Projections of error trajectories on normal planes

All three variables approach the origin of the co-ordinates with increasing time.

Transformation from one co-ordinate system to the other is given by equations 27 and 28.

$$\theta_s = \|M\| X_s \tag{27}$$

$$X_s = \|M\|^{-1} \theta_s \tag{28}$$

The modal matrix $\|M\|$ and its inverse may be determined by conventional methods.[9] The projections of the system trajectories on the normal planes $X_{s3}-X_{s1}$ and $X_{s3}-X_{s2}$, and the isoclines of these projections are obtained from equations 21 to 23.

$$X_{s3} = \frac{X_{s30}}{(X_{s10})^{\lambda_3/\lambda_1}} (X_{s1})^{\lambda_3/\lambda_1} \tag{29}$$

$$X_{s3} = \frac{X_{s30}}{(X_{s10})^{\lambda_3/\lambda_1}} (X_{s1})^{\lambda_3/\lambda_1} \tag{30}$$

$$X_{s3} = \frac{\lambda_3}{\lambda_2} \beta_{31} X_{s1}, \text{ where } \beta_{31} = \frac{dX_{s3}}{dX_{s1}} \tag{31}$$

$$X_{s3} = \frac{\lambda_3}{\lambda_2} \beta_{31} X_{s1}, \text{ where } \beta_{31} = \frac{dX_{s3}}{dX_{s1}} \tag{32}$$

The projections of the trajectories are plotted easiest by the method of isoclines (equations 31 and 32). Typical projections are shown in Fig. 9. The arrows indicate the directions of increasing time.

Positive Feedback

The system may be described in terms of the actuating signal (a, e, and c being positive constants)

$$\ddot{\theta}_s + a\ddot{\theta}_s + e\dot{\theta}_s - c\theta_s = 0 \tag{33}$$

or in terms of the error

$$\ddot{\theta}_e + a\ddot{\theta}_e + e\dot{\theta}_e - c\theta_e + 2c\theta_i = 0 \tag{34}$$

Equation 33 is independent of the input signal magnitude; equation 34 is not. The same applies to the respective trajectories.

To determine the desired switching mode (positive or negative feedback), the trajectories of both modes are superimposed on one set of reference co-ordinates. It is most practical to use the error trajectories.

By detailed analysis of the normal co-ordinate transformation, it is found that, in the normal co-ordinate system of the positive-feedback trajectories, the error trajectories, for any given input magnitude, may be obtained from the actuating-signal trajectories by translating the co-ordinates to point

$$\left(-\frac{2uc\theta_i}{\delta_1}, -\frac{2vc\theta_i}{\delta_2}, -\frac{2wc\theta_i}{\delta_3}\right)$$

where u, v, and w are, respectively, the terms of the third column, first, second, and third row, of the inverse modal matrix; c is the coefficient of the error term in the differential equation of the system; δ_1, δ_2, and δ_3 are the roots of the characteristic equation of the system, and θ_i is the input signal.

Transformation of the trajectories to their normal co-ordinates and translation along the axes as indicated above, yield the following equations for the normal projections of the error trajectories.

$$\left(X_{es2} - \frac{2vc\theta_i}{\delta_2}\right) = C_1 \left(X_{es1} - \frac{2uc\theta_i}{\delta_1}\right)^{\delta_2/\delta_1} \tag{35}$$

$$\left(X_{es3} - \frac{2wc\theta_i}{\delta_3}\right) = C_2 \left(X_{es1} - \frac{2uc\theta_i}{\delta_1}\right)^{\delta_3/\delta_1} \tag{36}$$

where

$$C_1 = \frac{\left(X_{es20} - \frac{2vc\theta_i}{\delta_2}\right)}{\left(X_{es10} - \frac{2uc\theta_i}{\delta_1}\right)^{\delta_2/\delta_1}}$$

$$C_2 = \frac{\left(X_{es30} - \frac{2wc\theta_i}{\delta_3}\right)}{\left(X_{es10} - \frac{2uc\theta_i}{\delta_1}\right)^{\delta_3/\delta_1}}$$

A set of typical actuating-signal trajectories, obtained by the method of isoclines, is shown in Fig. 10. Part of the trajectories have been transformed into error trajectories (with a change in scale), by translation of the normal co-ordinates, as shown in Fig. 11. The arrows indicate directions of increasing time, as determined from the three independent equations of the variables in the normal space, in analogy to the negative-feedback case.

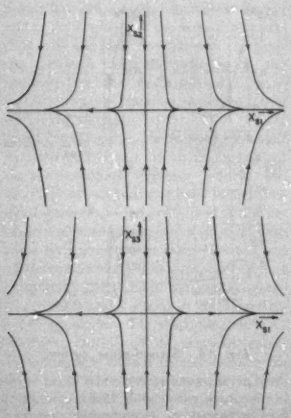


Fig. 10. Third-order positive-feedback system. Projections of actuating signal trajectories on normal planes

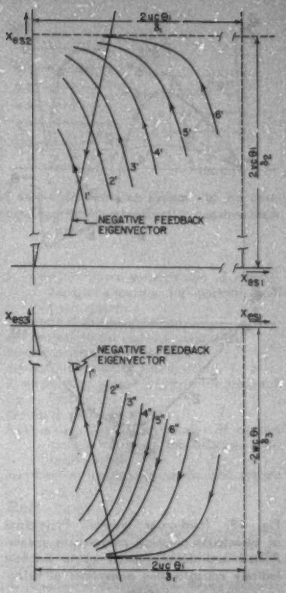


Fig. 11. Third-order positive-feedback system. Projections of error trajectories on normal planes. Primed and double-primed arabic numbers indicate corresponding projections of trajectories which intersect the negative-feedback eigenvector

The transformations relating the θ reference system and the X reference system are

$$\theta_e = \|N\| X_{es} \tag{37}$$

$$X_{es} = \|N\|^{-1} \theta_e \tag{38}$$

$\|N\|$ is the modal matrix of the set of simultaneous equations corresponding to equation 33.

Transformations between the negative-feedback normal space and the positive-feedback normal space are obtained by proper partial combinations of equations 27, 28, 37, and 38.

Determination of Operation Modes

The general reasoning is as follows: To have a stable system, the system has to approach the final state along a negative-feedback trajectory. An accepted criterion for the establishment of the response time of the type of systems under discussion, requires that the system reach its final

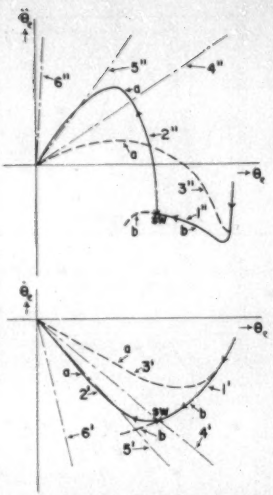


Fig. 12. Third-order system. Projections of trajectories on main phase-space planes. Primed and double-primed arabic numbers indicate corresponding projections of trajectories

a = negative feedback
b = positive feedback
1,2 = trajectories traversed by switching-type system
3 = trajectory transversed by linear negative-feedback system
4,5,6 = eigenvectors of negative-feedback system
sw = switching point

position with no overshoot. This may be secured by having the system approach the final state along an eigenvector of the negative-feedback trajectories. Switching along the eigenvector also provides a simple linear switching condition.

The eigenvectors are determined by the characteristic roots, λ. Since it was assumed that all λ's were distinct, there are three eigenvectors. The equations of the eigenvectors are

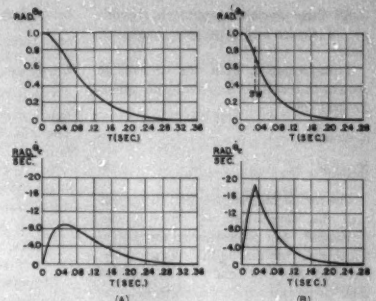
$$\dot{\theta}_e = \lambda_n \theta_e \tag{39}$$

$$\ddot{\theta}_e = \lambda_n^2 \theta_e \tag{40}$$

An eigenvector may be reached only from those positive trajectories which form a surface that includes the particular eigenvector. This surface, in turn, may be reached from negative trajectories. Which of the three surfaces is reached depends on the initial conditions of the system.

Fig. 13 (right). Second-order systems

A—Oscillogram of second-order, critically damped, linear negative-feedback system
B—Oscillogram of second-order switching system with no overshoot. sw = switching instant from positive feedback to negative feedback



The fastest eigenvector is the one the projection of which on the $\dot{\theta}_e - \theta_e$ plane has the smallest (most negative) slope with respect to the θ_e axis; i.e., the one which corresponds to the largest (absolute value) characteristic root.

The positive-feedback switching surfaces may now be obtained. This will be explained for the surface which corresponds to the eigenvector given, say, by the equations $\dot{\theta}_e = \lambda_1 \theta_e$ and $\ddot{\theta}_e = \lambda_1^2 \theta_e$.

The projections of the eigenvector in the positive-feedback normal phase space are obtained by transformation 38.

$$X_{es1} = pX_{es1} \tag{41}$$

$$\dot{X}_{es1} = qX_{es1} \tag{42}$$

where p and q are constants.

The positive-feedback switching surface is composed of all trajectories which intersect the eigenvector. This ensemble of trajectories is characterized by a specific functional relationship between the parameters C_1 and C_2 of equations 35 and 36. Combination of equations 35 and 39 determines the relationship between C_1 and X_{es1} at the point of intersection. Similarly, combination of equations 36 and 40 determines the relationship between C_2 and X_{es1} at the point of intersection. Elimination of X_{es1} from these last two expressions, determines the required relationship between C_1 and C_2. X_{es1} cannot be eliminated directly. C_1 and C_2 are plotted against each other point by point. The required function is then obtained by a convenient curve-fitting method to any desired accuracy. The eigenvector and positive-feedback switching trajectories are shown in Fig. 11. (The switching trajectories are indicated by corresponding primed and double-primed arabic numbers.) The range of interest of values for X_{es1} is generally limited and may be determined from Fig. 11. It should be noted that the curve fitting (which determines the first switching instant) is not critical, if the condition for the second switching point along the eigenvector is widened to include a cylindrical volume about the eigenvector. In some cases it is sufficient to express the first switching condition, in the region of interest, by a linear relation.

$$C_2 = mC_1 + Z \tag{43}$$

The switching criterion given above requires two switching actions.

The initial state may be such that the switching surface is never intersected. Such case is represented in Fig. 12. Advantage of positive feedback to shorten response time may be taken by starting the response along a positive-feedback trajectory and switching into negative feedback when a predetermined trajectory (or any of a group of trajectories)

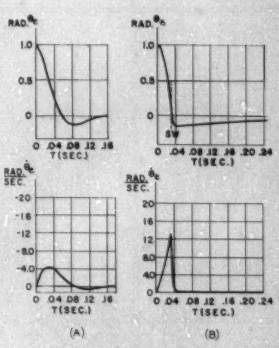


Fig. 14. Second-order systems

A—Oscillogram of second-order linear negative-feedback system with 12.5% overshoot
B—Oscillogram of second-order switching system with 12.5% overshoot (for unit step input). sw = switching instant from positive feedback to negative feedback

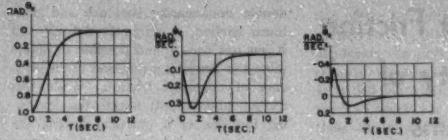


Fig. 15. Oscillogram of third-order linear negative-feedback system

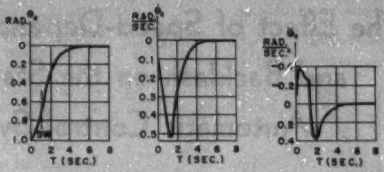


Fig. 16. Oscillogram of third-order switching system with no overshoot. sw = switching instant from positive feedback to negative feedback

is reached. This requires only one switching action.

The response may be shortened further by varying the damping co-efficient as pointed out in the section on second-order systems.

Note

The normal co-ordinate transformation was used for the general analysis. For a specific system with a small number of possible initial conditions it might be possible to determine the switching conditions directly from the projections of the few possible trajectories on the major planes of the main θ_s space. These trajectories may be plotted point by point by the use of a digital computer.

Experiments

All the experiments were simulated on an analog computer.

Second-Order System

Fig. 13 shows: 1. the response to a unit position step input of a linear negative-feedback system represented by the differential equation with initial conditions: $\ddot{\theta}_s+4\dot{\theta}_s+4\theta_s=0$; $\theta_{s0}=1$, $\dot{\theta}_{s0}=0$; and 2. the response of the same system except that the response starts in the positive-feedback mode and switches into the negative-feedback mode when $\dot{\theta}_s+2\theta_s=0$. The response time is reduced to 0.6 of its previous value.

Fig. 14 shows: 1. the response to a unit position step input of an underdamped negative-feedback system which has 12.5% overshoot, and where initial conditions were $\theta_{s0}=1$, $\dot{\theta}_{s0}=0$; and 2. the response of a switching-servo system with same overshoot under same external and initial conditions. The damping factor of the system is $\zeta=0.5$ in the positive-feedback mode and $\zeta=10.0$ in the negative-feedback mode. The response time is reduced to 0.68% of its previous value. (The response time is defined here as the time required to reduce the error to zero, before the system overshoots.)

Third-Order System

Fig. 15 shows the response to a unit position step input of a purely negative-feedback system represented by the differential equation

$$\dddot{\theta}_s+6\ddot{\theta}_s+8.85\dot{\theta}_s+3.75\theta_s=0$$

with the following initial conditions:

$$\theta_{s0}=1,\ \dot{\theta}_{s0}=-0.1,\ \ddot{\theta}_{s0}=-0.1$$

Fig. 16 shows the response of the same system of the preceding case with the same initial conditions, except that the system starts to respond with a positive-feedback connection, and switches into negative feedback under a predetermined condition. The response time is shortened to 0.63% of previous case.

Conclusions

Direct positive-negative feedback-switching servo systems prove valuable in particular for step and jump position input functions. The switching conditions vary from very simple to complicated ones as the possible states of the system increase. The system may be designed economically if the possible input magnitudes and initial conditions are known. Switching along an eigenvector provides a linear switching condition.

Nomenclature

s = Laplace operator
t = time
X_{sn} ($n=1, 2, 3$) = n's co-ordinate of positive-feedback actuating signal normal phase space
X_{en} ($n=1, 2, 3$) = n's co-ordinate of negative-feedback error normal-phase space
X_{stn} ($n=1, 2, 3$) = n's co-ordinate of positive-feedback error normal-phase space
α = slope of trajectory (second-order system)
α_e = eigenvalue (second-order characteristic equation)
β = slope of trajectory (third-order system)
β_n ($n=1, 2, 3$) = n's eigenvalue of characteristic equation for positive-feedback system
ζ = relative damping factor
θ_x = critical signal causing saturation
θ_i = input signal of the system
θ_0 = output signal of the system
θ_s = actuating signal of positive-feedback system
θ_e = error signal of system
λ_n ($n=1, 2, 3$) = n's eigenvalue of characteristic equation for negative-feedback system
ω_n = natural frequency of undamped system

References

1. Nonlinear Techniques for Improving Servo Performance, D. McDonald. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 400–21.

2. The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism, J. B. Lewis. *AIEE Transactions*, vol. 71, pt. II, 1952 (Jan. 1953 section), pp. 449–53.

3. A Phase-Plane Approach to the Compensation of Saturating Servomechanisms, A. M. Hopkin. *Ibid.*, vol. 70, pt. I, 1951, pp. 631–39.

4. An Investigation of the Switching Criteria for Higher Order Contactor Servomechanisms, Irving Bogner, Louis F. Kazda. *Ibid.*, vol. 73, pt. II, July 1954, pp. 118–27.

5. On a Nonlinear Transfer System, I. Flügge-Lotz, W. S. Wunch. *Journal of Applied Physics*, New York, N. Y., vol. 26, no. 4, 1955, pp. 484–88.

6. Synthesis of a Nonlinear Control System, I. Flügge-Lotz, C. F. Taylor. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., May 1956, pp. 5–9.

7. Theory of Oscillations (book), A. A. Andronow, C. E. Chaikin. Princeton University Press, Princeton, N. J., 1949.

8. Automatic Feedback Control System Synthesis (book), J. G. Truxal. McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

9. The Mathematics of Circuit Analysis (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1949.

10. Stability Theory of Differential Equations (book), R. Bellman. McGraw-Hill Book Company, Inc., 1953.

11. Lectures on Differential Equations (book), S. Lefschetz. Princeton University Press, 1946.

12. Analysis of Nonlinear Systems with More Than One Degree of Freedom by Means of Space Trajectories, Y. H. Ku. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 259, no. 2, 1955, pp. 115–31.

13. Analysis and Design Principles of Second and Higher Order Saturating Servomechanisms, R. E. Kalman. *AIEE Transactions*, vol. 74, pt. II, Nov. 1955, pp. 294–310.

The Effect of Speed-Dependent Friction and Backlash on the Stability of Automatic Control Systems

E. A. FREEMAN
NONMEMBER AIEE

Synopsis: A control system is considered which employs a motor driving a load through gearing having backlash. It is assumed that linear speed-dependent friction is applied to both the motor and the load. The effect of backlash and friction on the stability of the control system is investigated using the describing function technique. A method of deriving the describing function is given and it is shown that the presence of friction modifies the application of the describing function as presented by Tustin.[1] In particular it is shown that although the positional reset may be divided in the ratio of the inertias, the system may still exhibit sustained oscillations, and the choice of system parameters to eliminate these oscillations is discussed. To illustrate the effect of backlash and friction quantitatively a second-order control system is considered. The oscillations predicted by the theory are measured on an electronic analog of the system. By a correct choice of the system parameters the oscillations are removed.

IN THE PRODUCTION of remote position control systems there is a limit to the accuracy with which the motor and load can be connected. Generally the motor and load are connected through gearing which, to be practical, must possess some backlash. The effect of this backlash on the performance of the control system may, in certain cases, be of little importance. For example, if the magnitude of the backlash is small compared with the accuracy limit of the system any limit cycle occurring due to backlash may be insignificant. Further, the transient in response to an input whose amplitude is large compared with the backlash will be little different from the response of the system in the absence of backlash. However, as the restrictions on accuracy of control become more stringent, the effect of backlash on control system performance assumes a greater importance.

The introduction of the describing function technique, for the determination of the stability of nonlinear control systems, has made it possible to determine the effect of backlash on stability. A difficulty arising in applying this technique is the calculation of the describing function. The problem is to determine the fundamental component of the output from the nonlinearity in terms of the amplitude of the input sinusoid. In the case of simple, single-valued nonlinearities,[2] the fundamental component of the output can be calculated without knowing the harmonics present in the output. When backlash is considered, the interdependence of the fundamental and harmonic components of the output prevents the exact calculation of the fundamental component describing function without a knowledge of the harmonic content of the output.

Previous studies[1,3] of this problem have overcome this difficulty by assuming that the center of mass of the motor and load moves sinusoidally when the torque developed by the motor is sinusoidal. The practical application of the describing function technique using this assumption has provided much information concerning the effect of backlash. However, the theory fails to predict the effect of divided reset which, it is suggested, can be chosen so as to obtain a response as stable as would be obtained with the linear system. An alternative approach[4] assumes that the difference between the motor and load position is sinusoidal for a sinusoidal motor torque. The work of Liversidge[5] and the analysis presented in this paper have shown that it is a reasonable approximation to assume that the center of mass moves sinusoidally. It follows that the assumption that the difference between motor and load velocity is sinusoidal is not justified.

It is clear that any analysis using the describing function technique to determine absolute stability should use the fundamental component describing function calculated when harmonics are present in the output. These harmonics are caused by 1. the discontinuity in the relationship between motor and load position, 2. the presence of linear speed-dependent friction, and 3. the presence of nonlinear friction and stiction.

The analysis of a system possessing the double nonlinearity backlash and nonlinear friction, is extremely difficult and it has not been found possible to determine the describing function in the presence of nonlinear friction. While this friction undoubtedly affects the describing function of a practical system, the analysis presented in this paper focuses attention on the effect of speed-dependent friction on the nonlinearity backlash. The method presented for determining the fundamental component describing function neglects harmonics. A correction factor due to the presence of harmonics is then derived. It is found that the presence of speed-dependent friction can lead to an extra phase shift in the position of the center of mass not accounted for by previous theories. This phase shift can lead to the production of small-amplitude low-frequency oscillations which are associated with backlash in practical control systems. By a suitable choice of divided reset these oscillations can be eliminated.

Application of the Describing Function Technique to Backlash

The control system to be considered is assumed to have a motor as the power element driving a load through gearing having backlash. The effect of backlash and speed-dependent friction on the stability of this control system is to be examined using the describing function technique.[6,7] In order to apply the describing function to the nonlinearity backlash it is neecessary to determine the amplitude and phase of the fundamental component of the output, when the input waveform is sinusoidal. The input to the nonlinearity is defined as the torque developed by the motor. The output from the nonlinearity is regarded as the load position when positional reset is taken from the load and the motor position when positional reset is taken from the motor.

The following method is used to calculate the amplitude and phase of the

Paper 58-1268, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted December 16, 1957; made available for printing August 29, 1958.

E. A. FREEMAN is with The English Electric Company, Ltd., Whetstone, England.

The author is indebted to Prof. J. C. Prescott of King's College, Newcastle upon Tyne, for the facilities placed at his disposal, and to Dr. C. A. Walley for his helpful criticism and encouragement during the course of the work described. Thanks are also due to F. J. U. Ritson for his advice and encouragement and to Dr. M. Hutton, head of the Mathematics Department, The Technical College, Sunderland, for helpful discussions on the mathematical approach to the problem.

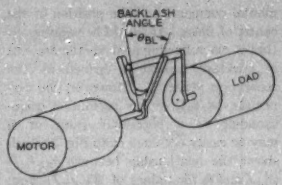


Fig. 1. Schematic representation of motor and load connected through gearing having backlash

fundamental component of the output. Firstly, the amplitude and phase of the fundamental component of the velocity of the center of mass is determined. Secondly, the amplitude and phase of the fundamental component of the velocity of the output relative to the velocity of the center of mass is determined. The amplitude and phase of the output position then follows by integration.

The velocity of the center of mass $\dot{\theta}$ is defined by equation 1.

$$J_M\dot{\theta}_M+J_L\dot{\theta}_L=(J_M+L_L)\dot{\theta}=J\dot{\theta} \qquad (1)$$

where

J_M = moment of inertia on motor side of backlash
J_L = moment of inertia on load side of backlash
$\dot{\theta}_M$ and $\dot{\theta}_L$ = the motor and load velocities, respectively

In the absence of friction it may be shown that the center of mass moves sinusoidally with the same amplitude and phase as in the absence of backlash.[7] However, due to friction, which is inevitably present in any practical system, the amplitude and phase of the fundamental component of the velocity of the center of mass are modified. Later the modifications which take place due to friction are investigated.

Motion of the Motor and Load for a Sinusoidal Motor Torque

Consider the motor and load connected through gearing having backlash. The schematic arrangement of the motor and load is shown in Fig. 1.

Assume that the velocity of the center of mass is a sinusoidal function of time when the torque developed by the motor is a sinusoidal function of time. (The justification for this assumption is given later.) When the motor and load are connected through the gearing such that $\dot{\theta}_M=\dot{\theta}_L$, it is seen from equation 1 that $\dot{\theta}=\dot{\theta}_M=\dot{\theta}_L$ (see Fig. 2); the force exerted on the load, through the gearing, is just sufficient to make the acceleration of the load equal to the acceleration of the motor. When the motor starts to decelerate the only force tending to decelerate the load is the frictional force. Provided the frictional force on the load tends to decelerate the load more quickly than the motor torque decelerates the motor, the load will maintain contact with the motor. At the instant for which the deceleration of the load becomes less than the deceleration of the motor the load and motor become separated (Fig. 3). During the separation period the velocity of the load is determined by the frictional forces acting.

At any instant during the separation the distance moved by the motor relative to the load is

$$\theta_M-\theta_L=\int(\dot{\theta}_M-\dot{\theta}_L)\,dt \qquad (2)$$

When this relative distance becomes equal to the backlash angle θ_{BL} contact is re-established. It can be seen that the separation period θ_0/ω ends when the shaded area is equal to the backlash angle, θ_{BL}; see Fig. 2.

It is assumed that there are no losses at the instant of making contact and that the motor and load move on together after the impact. Furthermore, it is assumed that the natural frequency of the load[4] is very much greater than any frequency of oscillation to be considered.

Amplitude and Phase of the Center of Mass Velocity with Viscous Friction

Consider the equations defining the system:

When motor and load are connected through the gearing

$$T_q=J_M\ddot{\theta}_M+K_{VM}\dot{\theta}_M+J_L\ddot{\theta}_L+K_{VL}\dot{\theta}_L \qquad (3)$$

When motor and load become separated

$$T_q=J_M\ddot{\theta}_M+K_{VM}\dot{\theta}_M \qquad (4)$$

$$0=J_L\ddot{\theta}_L+K_{VL}\dot{\theta}_L \qquad (5)$$

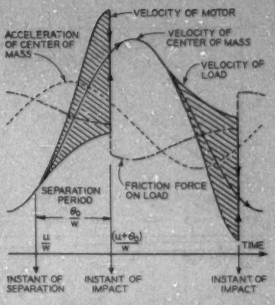


Fig. 2. Waveforms of motor velocity and load velocity when the velocity of the center of mass is sinusoidal

Adding equations 4 and 5,

$$T_q=J_M\ddot{\theta}_M+K_{VM}\dot{\theta}_M+J\ddot{\theta}_L+K_{VL}\dot{\theta}_L \qquad (6)$$

Comparing equation 3 with equation 6 it is seen that one equation defines the motion during contact and separation. Substituting α and writing $D=d/dt$, the equation of motion becomes:

$$T_q=J\left\{D+K_{VM}/J_M\right\}\dot{\theta}-\frac{1-\alpha}{\alpha}K_{VL}\dot{\theta}_L \qquad (7)$$

If the motor torque T_q is harmonic-free, the harmonics of $J(D+K_{VM}/J_M)\dot{\theta}$ must vanish identically with the harmonics of $(1-\alpha/\alpha)K_{VL}\dot{\theta}_L$. Since $\dot{\theta}_L$ possess harmonics, $\dot{\theta}$ must also possess harmonics. These harmonics of $\dot{\theta}$ affect the fundamental component of $\dot{\theta}_L$ which in turn modifies the fundamental component of $\dot{\theta}$. Because of this interdependence between the fundamental component and harmonics of $\dot{\theta}$ it follows that the fundamental component describing function cannot be calculated exactly unless the harmonic content of $\dot{\theta}$ is known. By using an iteration process (Appendix I)

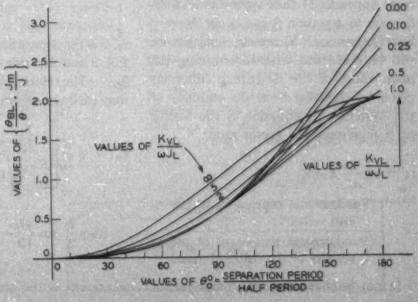


Fig. 3. Relationship between the separation period θ_0/ω and the backlash angle θ_{BL}

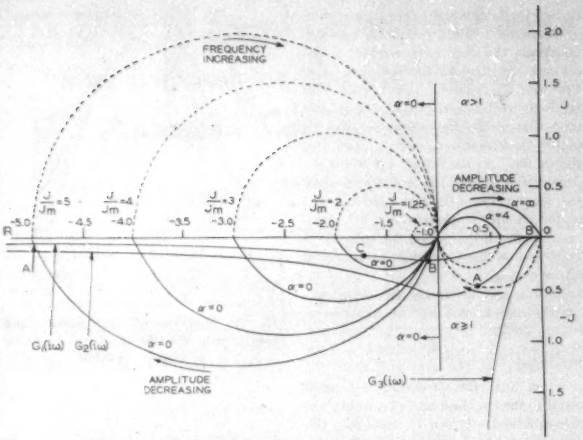


Fig. 4. Envelope of loci $(-1/N_1)N_2$ is describing function which gives effect of friction and backlash on the velocity of the center of mass

——— Zero frequency value of $-1/N_1$
– – – – – Locus of the extremities of the $-1/N_1$ loci for variable frequency
The curves $G_{1,2,3}(j\omega)$ represent the open-loop frequency response of control system

it may be shown that a first approximation to the fundamental component of $\dot{\theta}$ is given by:

$$\dot{\theta}=\frac{RT_\theta e^{-i\phi}}{\{1-rRe^{-i(\phi+\psi)}\}} \tag{8}$$

Equation 8 applies when the harmonics of $\dot{\theta}$ are small. At low frequencies, however, the harmonic content of $\dot{\theta}$ becomes important. The determination of the describing function at low frequency thus entails the calculation of the harmonics. In Appendix II the first correction factor to equation 8 is derived.

This correction factor takes account of the effect of the third harmonic on the fundamental component of $\dot{\theta}$. It follows from Appendix II that appreciable modification to equation 8 can occur at very low frequencies. However, initially we will be concerned with determining the main effect of the describing functions given by equation 8 on the stability of the system. In equation 8 the vector $re^{-i\psi}$ represents the vector ratio:

$$re^{-i\psi}=\frac{1-\alpha}{\alpha}K_{VL}\times \frac{\text{(Fundamental component of the load velocity)}}{\text{(Fundamental component of the velocity of the center of mass)}}$$

It may be shown that the component of $\dot{\theta}_L$ in phase with $\dot{\theta}$ is $r\cos\psi=a$, where a is given by equation 9.

$$a=\frac{1}{\pi}\left[\pi-\theta_0+\sin\theta_0\times\cos\left\{2\tan^{-1}\left(\frac{\omega J_L}{K_{VL}}\right)-\theta_0\right\}\right]+\frac{2}{\pi}\frac{\epsilon^{-\frac{K_{VL}}{\omega J_L}\theta_0}\cos\theta_0}{\{1+(K_{VL}/\omega J_L)^2\}} \tag{9}$$

Similarly the component of $\dot{\theta}_L$ lagging $\dot{\theta}$ by 90 degrees is $r\sin\psi=b$ where b is given by equation 10.

$$b=\frac{1}{\pi}\sin\theta_0\sin\left\{2\tan^{-1}\left(\frac{\omega J_L}{K_{VL}}\right)-\theta_0\right\}-\frac{2}{\pi}\frac{\left\{1-\epsilon^{-\frac{K_{VL}}{\omega J_L}\theta_0}\cos\theta_0\right\}}{\{1+(K_{VL}/\omega J_L)^2\}} \tag{10}$$

θ_0/ω is the separation period as defined in Fig. 2 and depends on the backlash angle θ_{BL}. The relationship between θ_0 and θ_{BL} may easily be shown to be

$$\theta_{BL}=|\theta|\frac{J}{J_M}\left[\frac{\left\{1-\epsilon^{-\frac{K_{VL}}{\omega J_L}\theta_0}\right\}}{\frac{K_{VL}}{\omega J_L}\sqrt{1+\left(\frac{K_{VL}}{\omega J_L}\right)^2}}+2\sin\frac{\theta_0}{2}\sin\left\{\frac{\theta_0}{2}-\tan^{-1}\left(\frac{\omega J_L}{K_{VL}}\right)\right\}\right] \tag{11}$$

where $|\theta|$ is the amplitude of the fundamental component of the position of the center of mass. It should be pointed out that in the following analysis the amount of backlash present is defined by the separation period θ_0. However, the amplitude, $|\theta|/\theta_{BL}$, of any oscillation corresponding to a particular value of θ_0 may be easily obtained from Fig. 3 which shows the relationship between $\theta_{BL}J_M/|\theta|\ J$ and θ_0 for values of $K_{VL}/\omega J_L$ from zero to infinity.

Describing Functions Showing the Effect of Backlash and Friction on the Velocity of the Center of Mass

In the absence of backlash the fundamental component of $\dot{\theta}$ may be shown to be

$$\dot{\theta}=\frac{RT_\theta e^{-i\phi}}{\{1-r'Re^{-i\phi}\}} \tag{12}$$

Defining the describing function N_1 by equation 13

$$N_1=\frac{\dot{\theta}\text{ when backlash is present}}{\dot{\theta}\text{ when backlash is absent}} \tag{13}$$

N_1 becomes:

$$N_1=\frac{\{1-r'Re^{-i\phi}\}}{\{1-rRe^{-i(\phi+\psi)}\}} \tag{14}$$

so that $-1/N_1$ becomes:

$$-\frac{1}{N_1}=\frac{-\{1-rRe^{-i(\phi+\psi)}\}}{\{1-r'Re^{-i\phi}\}} \tag{15}$$

The loci defined by equation 15 are both amplitude and frequency dependent. For any particular set of parameters these loci are contained within an envelope. The position of these envelopes in the complex plane is of particular importance to the stability of the system.

The sides of the envelope are defined by equations 16 and 17

$$-\frac{1}{N_1}\underset{\psi\to\pi/2}{=}-\frac{\{1-rR_4^{-i(\phi+\psi)}\}}{\{1-r'Re^{-i\phi}\}} \tag{16}$$

$$-\frac{1}{N_1}\underset{\omega\to 0}{=}-\frac{\{1-rRe^{-i(\phi+\psi)}\}}{\{1-r'Re^{-i\phi}\}} \tag{17}$$

It can be seen from the loci of Fig. 4 that considerable phase shift of the center of mass may be produced by speed-dependent friction and backlash. Later the effect of the describing functions contained within these envelopes, on the stability of three specific control systems, is discussed.

Amplitude and Phase of the Output Position Relative to the Position of the Center of Mass

It will be assumed that in general the positional reset is divided, it being made

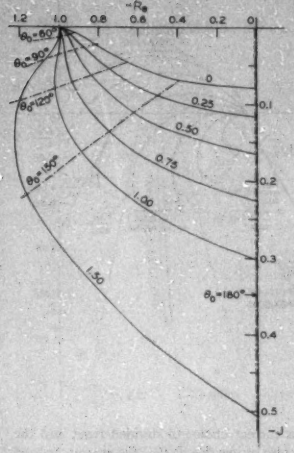


Fig. 5. Loci of $\frac{1}{a-ib}$ as function of θ_0 for values of $\frac{K_{VL}}{\omega J_L}$ in the range 0 to 1.5. The vector $(a-ib)$ represents the phase shift and amplitude change of θ_L relative to θ

----- Lines of constant θ_0

——— Lines of $\left(\frac{1}{a-ib}\right)$ for various values of $\frac{K_{VL}}{\omega J_L}$. The appropriate value of $\frac{K_{VL}}{\omega J_L}$ is marked on each curve

up of signals from motor position and load position. The stability of the system with divided reset can be examined by determining the stability of the output position as defined by equation 18

$$\theta' = k_1\theta_M + k_2\theta_L \tag{18}$$

and

$$k_1 + k_2 = 1$$

θ' being the output position.

The fraction of positional reset taken from the motor and load are given by k_1 and k_2, respectively. Since the maximum phase displacement between the fundamental components of θ_M and θ_L is less than π it follows that if a periodic form of θ' is found to exist, then this is a necessary and sufficient condition for the existence of periodic forms of θ_M and θ_L. Assuming that θ is harmonic-free the amplitude and phase of θ_L relative to θ may be calculated. One obtains:

$$\theta_L = \sqrt{a^2+b^2}\, e^{-i\psi}\theta \tag{19}$$

where a and b are defined by equations 9 and 10 and

$$\psi = \tan^{-1}(b/a)$$

= the phase shift of θ_L relative to θ

Defining a second describing function N_2' by

$$N_2' = \frac{\text{Fundamental component of } \theta_L}{\text{Fundamental component of } \theta} \tag{20}$$

gives

$$N_2' = \sqrt{a^2+b^2}\, e^{-i\psi}$$

However, it is necessary to know the describing function for the output position θ'. Defining N_3 by equation 20,

$$N_3 = \frac{\text{Fundamental component of } \theta'}{\text{Fundamental component of } \theta}$$

It follows from equation 18 that

$$\theta' = \frac{J}{J_M}\, k_1\theta + (1 - k_1 J_L/k_2 J_M)\theta_L$$

Hence

$$N_3 = \frac{J}{J_M}\, k_1 + (1 - k_1 J_L/k_2 J_M) k_2 N_2' \tag{21}$$

Equation 21 gives the amplitude change and phase shift of the output position relative to the position of the center of mass. When positional reset is taken from the load $k_2 = 1$, $k_1 = 0$ and N_3 becomes equal to N_2'. The locus $-1/N_2'$ is of particular importance to the stability of the system and will be discussed in more detail in the next section. Curves of $-1/N_2' = -1/a + ib$ as a function of θ_0 are shown in Figs. 5 and 6; the amplitude of oscillation corresponding to any value of θ_0 may be determined from Fig. 3.

The Stability of a Control System in the Presence of Backlash and Viscous Friction

The system considered is assumed to consist of a linear frequency-dependent element, whose transfer function is $F(S)$, followed by a motor and load connected through gearing with backlash. It is assumed that viscous friction is applied to both the motor and load.

The transfer function $F(S)$ includes any time lag present in the motor field circuit.

It follows from Fig. 7 that

$$\frac{\theta'}{\theta_1} = \frac{N_1 N_2 F(S)/JS(S+K_V/J)}{\{1 + N_1 N_2 F(S)/JS(S+K_V/J)\}}$$

Writing

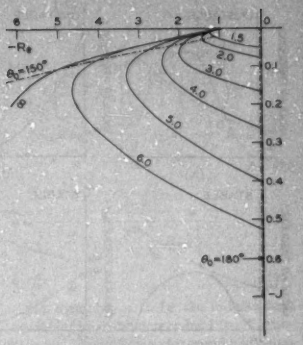


Fig. 6. Loci of $\frac{1}{a-ib}$ as a function of θ_0 for values of $\frac{K_{VL}}{\omega J_L}$ in the range of 1.5 to ∞. The vector $(a-ib)$ represents the phase shift and amplitude change of θ_L relative to θ

See Fig. 5 for curve identification

$$G(S) = \frac{F(S)}{JS(S+K_V/J)}$$

gives

$$\frac{\theta'}{\theta_1} = \frac{G(S)}{\{1/N_1N_2 + G(S)\}} \tag{22}$$

The loci $1/N_1$, $1/N_2$ are amplitude and frequency dependent, hence the condition for instability is

$$G(i\omega) = \frac{-1}{N_1 N_2} \tag{23}$$

Effect of the Describing Functions on the Stability of a Control System

In this section the effect of the describing functions N_1 and N_2 on the stability of a control system will be discussed.

It has been shown that the two main effects introduced by the presence of friction and backlash are

1. The fundamental component of the velocity of the center of mass undergoes a phase shift and amplitude change which are defined by the describing function N_1.

2. The fundamental component of the velocity of the load changes in amplitude and lags in phase relative to the velocity of the center of mass. This amplitude

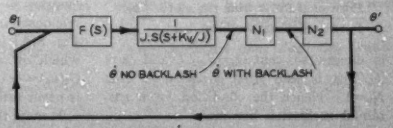


Fig. 7. Block diagram of control system with backlash

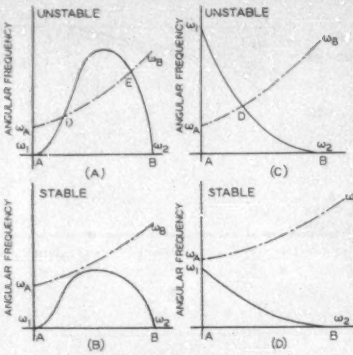


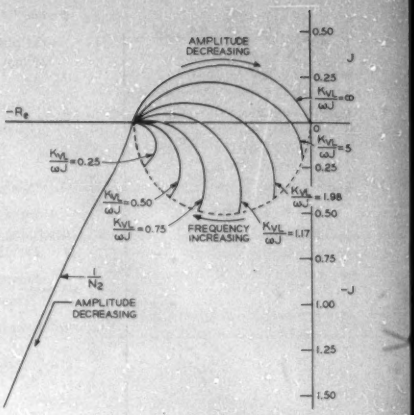
Fig. 8. Geometrical interpretation of the stability criterion with frequency-dependent describing functions

—— Frequency on $G(i\omega)$ locus

—— Frequency of the $-\frac{1}{N}$ locus being intersected

Distance AB represents that part of $G(i\omega)$ locus for which an intersection with $-\frac{1}{N}$ loci takes place

Fig. 9(right). Loci $-\frac{1}{N_1}$ which give modification to fundamental component of the velocity of the center of mass due to speed-dependent friction on load. Locus $-\frac{1}{N_2}$ gives the phase shift and amplitude change of θ_L relative to θ at very low frequencies



change and phase shift is represented by the describing function N_2.

In order to examine the effects in items 1 and 2 the parameters of the system will be chosen so that effect 2 is removed, effect 1 is removed, and finally effects 1 and 2 just described will be considered together.

For the purpose of discussing the stability of the system it is convenient to define certain angular frequencies:

Let $\omega_A/2\pi$ be the lowest frequency which causes $G(i\omega)$ to intersect a describing function.

Let the angular frequency corresponding to this describing function be ω_1.

Let $\omega_B/2\pi$ be the highest frequency which causes $G(i\omega)$ to intersect a describing function.

Let the angular frequency corresponding to this describing function be ω_2.

The Describing Functions N_1

When the ratio

$$\frac{\text{Positional Reset from the motor}}{\text{Positional Reset from the load}} = \frac{k_1}{k_2}$$

is chosen so that $k_1 = J_M/J$ and $k_2 = J_L/J$ it is seen from equation 21 that $N_2 = 1$. Hence the condition for a sustained oscillation becomes:

$$G(i\omega) = -1/N_1(\omega T_q)$$

The envelopes of the frequency-dependent loci $-1/N_1(\omega T_q)$ are shown in Fig. 4 together with three curves representing specific values of $G(i\omega)$. These curves will be referred to as the $G_{1,2,3}(i\omega)$ loci.

Consider the case where $J/J_M = 5$ and $\alpha = 0$ and assume that $G_1(i\omega)$ is the open-loop frequency response of the system. In moving through the envelope of the describing function from A to B the angular frequency of $G_1(i\omega)$ changes from ω_A to ω_B. The frequency of the $-1/N_1$ loci cut by the $G_1(i\omega)$ locus starts from zero at A, rises to a maximum at some point C, and then falls to zero at B. These conditions are shown in Fig. 8(A). At the points of intersection D and E, $G_1(i\omega) = -1/N_1(\omega T_q)$ at a common frequency. Hence oscillations are predicted at the points D and E. However, the oscillation corresponding to the point D may be shown to be unstable so that the system is predicted to exhibit sustained oscillations the amplitude and frequency of which are determined by the point E.

Consider now the effect of changing the ratio of the inertias J_L/J_M.

Let J/J_M be changed to 1.25.

In this case the $G_1(i\omega)$ locus does not move into the envelope of the describing functions and hence no oscillations are predicted.

Consider the effect of increasing the gain and phase margins of the linear systems so that the open-loop frequency response becomes $G_3(i\omega)$. In this case Fig. 8(A) is modified to Fig. 8(B) from which it may be seen that although $G_3(i\omega) = -1/N_1$ the frequency condition is never satisfied.

It may, therefore, be concluded that by a correct choice of divided reset and the ratio of inertias J_L/J_M the system defined by $G_1(i\omega)$ may be designed so that sustained oscillations of the output do not occur.

Change of the Ratio α

The ratio $\alpha = K_{VL}J_M/K_{VM}J_L$ also affects the stability of the system. Let α be changed to ∞; that is, assume that there is no viscous friction applied to the motor. From Fig. 4 it will be seen that as the $G_2(i\omega)$ locus moves from A to B the frequency of the describing functions being intersected decreases; see Fig. 8(C).

It follows, therefore, that if $\omega_A < \omega_1$ a sustained oscillation occurs. If $\omega_B > \omega_1$, no oscillation is predicted; see Fig. 8(D). It should also be pointed out that for systems with $G(i\omega)$ loci of the form $G_{2,3}(i\omega)$ it is possible to choose a value of α so that the open-loop frequency response does not enter the describing function envelope.

The most significant conclusion to be drawn from this discussion is that although the reset may be divided into the ratio of the inertias the stability criterion of the system is not the same as for the linear system.[1] This statement is in agreement with the experimental findings of Liversidge.[3]

The Describing Functions N_3

When $K_{VL}/J_L = K_{VM}/J_M$ the describing function N_1 becomes equal to unity.

In general the stability of the system is determined by the loci $-1/N_3'$. In particular when all positional reset is taken from the load the describing functions determining stability become N_3 (loci $-1/N_3$, Figs. 5 and 6). Referring to Figs. 4 and 5 it can be seen that all three

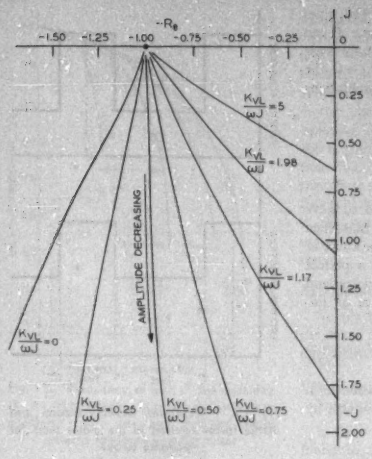


Fig. 10. Frequency-dependent describing functions $-1/N_1 N_2'$ used for examining the stability of second-order control system when positional reset is taken from the load. Speed-dependent friction on load is used for stabilization

loci $G_{1,3,5}(i\omega)$ intersect every one of the frequency-dependent loci $-1/N_2$.

In each case $\omega_A > \omega_1$ since $\omega_1 = 0$ and $\omega_B < \omega_3$ since $\omega_3 = \infty$, hence in all three cases sustained oscillations are predicted. This may be deduced by considering Fig. 8(A).

The Describing Functions $N_1 N_2'$

The number of parameters involved in the determination of the describing function $N_1 N_2'$ makes it very difficult to discuss the effect of the loci $-1/N_1 N_2'$ on stability in any general way. However, it can be shown that if all positional reset is taken from the load, the ratio of $K_{VL} J_M / K_{VM} J_L$ must be made less than unity if sustained oscillations are to be avoided with $aG(i\omega) = G_3(i\omega)$.

It should be pointed out that in order to prevent sustained oscillations with a $G(i\omega)$ of the form $G_{1,3}(i\omega)$ it may also be necessary to allow positional error and take some positional reset from the motor.

Examination of the Stability of a Second-Order System with Backlash Present

All Reset Taken from the Load

In this case $N_2 = N_2'$. The loci $-1/N_2'$ and $-1/N_1$ are shown in Fig. 9. The loci determining stability are $-1/N_1 N_2'$; see Fig. 10.

If the extra phase shift of the center of mass is neglected the loci determining stability becomes $-1/N_2'$ and $G(i\omega)$. It may be deduced from Fig. 10 that under such conditions the system would be predicted to be stable for a $K_{VL}/2\omega_0 J = 0.536$. When the effect of friction and backlash is taken into account instability is predicted for all values of damping.

The determinaton of the amplitude and frequency of the oscillation, using the graphical describing-function technique, becomes tedious when the describing functions are frequency dependent. In the case of a second-order system the conditions for the existence of a limit cycle can be derived explicitly without much difficulty. These conditions are derived in Appendix III from which it follows that a limit cycle exists provided equations 24, 25, and 26 can be satisfied simultaneously.

$$b^2 = \frac{\left\{\frac{K_{VM}}{J_M} J - \frac{1-\alpha}{\alpha} aK_{VL}\right\}^2 \times \left\{\frac{K_M}{J_M} J - \frac{1-\gamma}{\gamma} aK_L\right\}}{\left[J\left\{\frac{1-\gamma}{\gamma} K_L\right\}^2 + \frac{1-\alpha}{\alpha}\left(\frac{1-\gamma}{\gamma}\right) \times (K_L K_{VL})\left\{\frac{1-\alpha}{\alpha} K_{VL}(a) - \frac{J}{J_M} K_{VM}\right\}\right]} \quad (24)$$

$$\omega = \frac{\frac{1-\gamma}{\gamma} K_L b}{\left\{\frac{1-\alpha}{\alpha} K_{VL}(a) - \frac{J}{J_M} K_{VM}\right\}} \quad (25)$$

$$b = f(a) \quad (26)$$

By assuming a to be the independent variable, b can be determined from equation 26. Substituting for a and b in equation 24 gives K_{VL}. Hence ω can be calculated from equation 25.

From these calculations it has been found that an oscillation exists for all values of $0 \leq K_{VL} \leq \infty$. Figs. 11 and 12 show that as K_{VL} is increased the amplitude and frequency of the oscillation decrease.

All Reset Taken from the Motor

In this case equation 25 becomes (see Appendix II)

$$\omega = \left\{\frac{1-\gamma}{\gamma}(K_L) b\right\} \Big/ \left\{\frac{K_{VM}}{J_M} J(a - J/J_L)\right\}$$

Since $0 \leq a \leq 1$ and $J/J_L \geq 1$, $(a - J/J_L) < 0$. This implies ω is negative. Hence no oscillations exist.

The disadvantage of this method of stabilization is that a positional error equal to the dead space may exist. To avoid this error positional reset must be taken from the load.

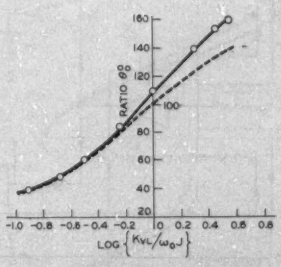


Fig. 11. Ratio θ, for the limit cycle occurring in a second-order system with speed-dependent friction on load and positional reset taken from load

- - - - Calculated using the describing-function technique
——— Exact calculation of the limit cycle
○ Measured value

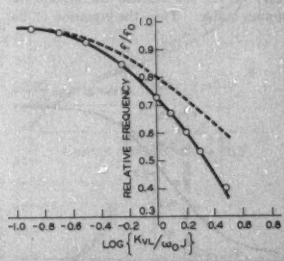


Fig. 12. Frequency of limit cycle occurring in a second-order system with speed-dependent friction on load and positional reset taken from load

- - - - Calculated using describing-function technique
——— Exact calculation of limit cycle
○ Measured value

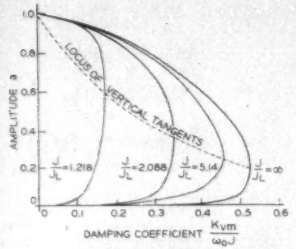


Fig. 13. Predicted amplitude of oscillation as a function of velocity feedback from motor for various values of inertia ratio. Positional reset from load

Positional Reset Taken from the Load, Velocity Signal Taken from the Motor

In this case equations 24 and 25 become (see Appendix III)

$$b^2=\frac{\{a(a-J/J_L)^2(K_V/\omega_0 J)^2\}}{\{(1-J/J_L)^2-(a-J/J_L)(K_V/\omega_0 J)^2\}} \tag{27}$$

$$\frac{3}{\omega_0}=\frac{\omega_0 J}{K_V}\cdot\frac{b}{J_L/J_M(J/J_L J-a)} \tag{28}$$

It follows from equations 26 and 27 that the curves of $(a\sim K_V/\omega_0 J)$ are not single valued for all $K_V/\omega_0 J$; see Fig. 13. This indicates two possible oscillations. However, by using the method of small increments,[8] it is possible to show that the lower value of a corresponds to an unstable oscillation. Hence only one oscillation occurs in a practical system. It follows that if $K_V/\omega_0 J$ is made greater than that value for which the $(a\sim K_V/\omega_0 J)$ curves are vertical the system becomes stable. Thus the equation of the vertical tangencies (Appendix IV) gives the stability limit for the system. It can be seen (Fig. 14) that in the case least favorable to stability (i.e., $J_L/J\to 1$) the value of $K_V/2\omega_0 J$ to prevent oscillations is 0.536.

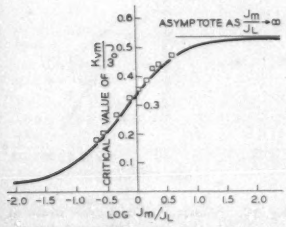


Fig. 14. Stability limit as function of the ratio of inertias. Graph of the values of $\frac{K_{VM}}{\omega_0 J}$ necessary to prevent sustained oscillations

☐ Measured values
——— Calculated values

Experimental

The Simulator

In the simulator it is wished to solve equations 3 and 4. It is assumed that when the motor has taken up the backlash at the end of a separation period a perfect impact occurs such that the motor and load move on together. Hence at impact it is necessary to solve:

Momentum before impact = Momentum after impact (29)

Equations 3 and 4 are of the same form and are thus solved in the simulator by solving the equation:

$$T_e=F(t)\ddot{\theta}_M+f(t)\dot{\theta}_M$$

where F and f are functions of time.

The variation of F and f is achieved by using diode switches which alter the gain of feedback amplifiers.

The functions F and f take the form shown in Fig. 15.

In order to determine the separation period an integrator is fed with a signal proportional to the difference between the motor and load velocity. This integrator is normally clamped, the clamp being released at the instant of separation.

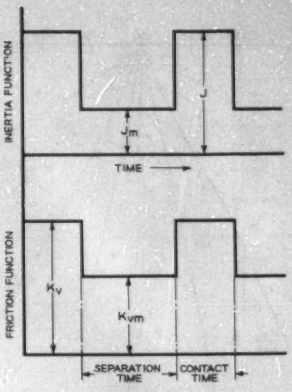


Fig. 15. Waveforms showing friction and inertia forces referred to the motor shaft for periodic torque

When the output of the integrator reaches a voltage corresponding to the backlash angle a pulse is derived defining the instant of contact and the integrator reset. The instant of separation is defined by a pulse derived at the instant when the difference between motor and load velocity becomes finite. The two pulses marking the instants of contact and separation are fed to a scale of two circuits, the output of which provides the basic switching

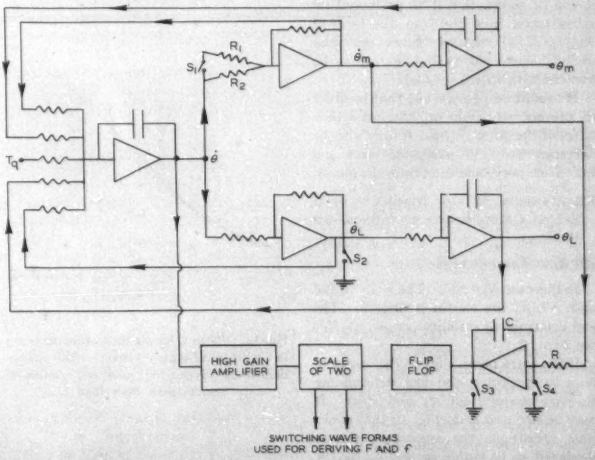


Fig. 16. Block diagram of analog for a system with combined backlash and viscous friction

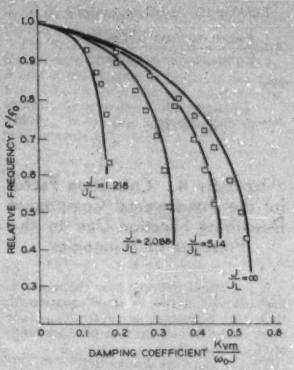


Fig. 17. Frequency of limit cycles occuring in a second-order control system when positional reset is taken from load and output velocity feedback from the motor

□ Measured values
——— Calculated values

waveforms producing F and f. (A block diagram of the analog of the system is shown in Fig. 16.)

Solution of the momentum equation 29 requires an instantaneous change in motor velocity at impact. The new velocity is given by equation 30.

$$\dot{\theta}_{M,A}=\frac{J_M\dot{\theta}_{M,B}+J_L\dot{\theta}_{L,B}}{J} \tag{30}$$

where

$\dot{\theta}_{M,A}$ = motor velocity after impact
$\dot{\theta}_{M,B}$ = motor velocity before impact

This is achieved by varying the gain of a feedback amplifier at the instant of contact.

Steady-State Oscillations

All Reset Taken from the Load

Oscillations were observed on the simulator for all values of damping in the range of $0.01 \leq K_{VL}/\omega_0 J \leq 4.1$. The frequency of the oscillation and the ratio θ_0 were measured for various values of damping. It can be seen from Figs. 11 and 12 that the calculated curves differ appreciably from those measured, the discrepancy increasing as the frequency of oscillation decreases. This difference between the calculation and measurement is due to the presence of harmonics in the motion of the center of mass which are not taken into account in the analysis. It has been shown (Appendix II) that a correction factor for the fundamental component describing function can be derived in terms of the first approximation. While this calculation can be applied to a system of any order a much more direct approach, which is possible in this simple case, is to determine the equation of the limit cycle. This involves joining the solutions of the equations defining the system in the two linear regions so as to produce a combined solution which is periodic. The calculation of the oscillation in this way leads to agreement within 2% with the measured results.

All Reset from the Motor

No oscillations were observed on the simulator although the damping was reduced almost to zero.

Positional Reset from the Load, Velocity Signal from the Motor

In this case the oscillations present were found to be dependent on the ratio of inertias. Further it was noted that as the damping was increased the oscillation became less stable until finally a value of damping was reached for which the oscillations vanished. This value of damping, called the stability limit, was measured for ratios of J_L/J_M in the range $0.242 \leq J_L/J_M \leq 4.59$. As can be seen from Fig. 14 good agreement with the analysis has been obtained.

The calculated curves showing frequency and the ratio θ_0 for fixed ratios of J_L/J_M and various damping coefficients (Figs. 17 and 18) differ appreciably from the measured values when the frequency is low, the effect being due to the presence of harmonics which are neglected in the analysis.

Conclusions

A method has been presented for determining the first approximation to the fundamental component describing function for backlash, when speed-dependent friction is present. It has been shown from this first approximation that the presence of speed-dependent friction can lead to an extra phase shift of the center of mass which, under conditions derived in the paper, may cause a control system to be less stable than existing theory predicts. In considering a second-order system with output velocity feedback, it has been found that, if all reset is taken from the load, the system becomes unstable for all values of damping coefficient. Stabilization has been achieved in two ways: 1. at the expense of positional error, by taking all reset from the motor; and 2. by taking positional reset from the load and the velocity signal from the motor. In this latter case a stability limit was predicted which agreed closely with measurements from the simulator.

In the two cases where oscillations were predicted, appreciable disagreement occurred between the theoretical and measured frequencies when the oscillation frequency was low. The reason for this disagreement is due to the presence of harmonics which, for this nonlinearity, have

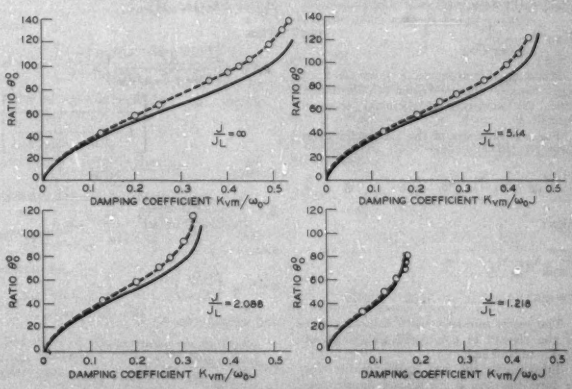


Fig. 18. Graph of θ_0 for the limit cycles occuring in a second-order control system, when positional reset is taken from the load and output velocity feedback from the motor

○ Measured values
——— Calculated using the describing function technique

large amplitude compared with the loop attenuation.

The inaccuracy of the first approximation has been avoided in the case considered by an exact calculation of the limit cycle. In general this calculation will not be possible. In order to improve the accuracy for any order system the correction factor, due to harmonics, for the first approximation has been derived. This correction factor has been determined in terms of the first approximation, so that an accurate calculation of the describing function is possible, once the waveform of load velocity is specified in terms of the center of mass velocity.

Appendix I. Determination of the Fundamental Component Describing Function

It has been shown that the equation determining the motion of the center of mass during contact and separation is

$$T_q = J\left(\ddot{\theta} + \frac{K_{VM}}{J_M}\dot{\theta}\right) - \frac{1-\alpha}{\alpha} K_{VL}\dot{\theta}_L \qquad (31)$$

It is desired to determine $\dot{\theta}$ when T_q is a sinusoidal function of time.

Let $T_q = A \sin \omega t = \mathcal{I}_e A^{i\omega t}$

It is assumed that the harmonics of $\dot{\theta}$ are negligible. A first approximation to $\dot{\theta}$ can then be obtained by assuming that:

$$\frac{1-\alpha}{\alpha}(K_{VL})\dot{\theta}_L = 0$$

Writing $\dot{\theta}_r$ for the rth approximation of $\dot{\theta}$ gives

$$\dot{\theta}_1 = \mathcal{I}\left\{\frac{A e^{i\omega t}}{J(D + K_{VM}/J_M)}\right\}$$

$$\dot{\theta}_1 = \mathcal{I}\left\{\frac{A e^{i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\} \qquad (32)$$

Knowing this aproximate value for $\dot{\theta}$ an approximate value for $\dot{\theta}_L$ can be determined from the nonlinear relationship between $\dot{\theta}_L$ and $\dot{\theta}$.

For a given value of the ratio-separation period/half-period time

$$\frac{1-\alpha}{\alpha}(K_{VL})\dot{\theta}_L = r e^{-i\psi}\dot{\theta}_1$$

where

$$\frac{1-\alpha}{\alpha} K_{VL} \sqrt{a^2+b^2} = r$$

$$\psi = \tan^{-1}(b/a)$$

The vector multiplier $re^{-i\psi}$ takes account of the change in amplitude and phase of $\dot{\theta}_L$ due to backlash.

Substituting the first approximation for $\dot{\theta}_L$ into equation 31, a second approximation to $\dot{\theta}$ may be made.

Thus

$$\mathcal{I} A e^{i\omega t} + r e^{-i\psi}\dot{\theta}_1 = J(D + K_{VM}/J_M)\dot{\theta}_2$$

Substituting for $\dot{\theta}_1$ from equation 32,

$$\dot{\theta}_2 = \mathcal{I} A e^{i\omega t} + r e^{-i\psi}\mathcal{I}\left\{\frac{A e^{i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\}$$

$$= \mathcal{I}\left\{\frac{A e^{i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\} \times \left\{1 + \frac{r e^{-i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\}$$

Using this second approximation for $\dot{\theta}$, the second approximation for $(1-\alpha/\alpha)K_{VL}\dot{\theta}_L$ can be calculated.

Substituting $(1-\alpha/\alpha)K_{VL}\dot{\theta}_{L2}$ in equation 31 results in

$$\dot{\theta}_3 = \mathcal{I}\left\{\frac{A e^{i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\} \times \left[1 + \left\{\frac{r e^{-i\psi}}{J(i\omega + K_{VM}/\omega J_M)}\right\} + \left\{\frac{r e^{-i\psi}}{J(i\omega + K_{VM}/J_M)}\right\}^2\right]$$

By repeating this process of successive approximation it may be shown that

$$\dot{\theta}_n = \mathcal{I}\left\{\frac{A e^{i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\} \times \left[1 + \left\{\frac{r e^{-i\psi}}{J(i\omega + K_{VM}/J_M)}\right\} + \cdots \left\{\frac{r e^{-i\psi}}{J(i\omega + K_{VM}/J_M)}\right\}^{n-1}\right]$$

Hence

$$\dot{\theta}_n = \mathcal{I}\left\{\frac{A e^{i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\} \times \left[\frac{1 - \left\{\frac{r e^{-i\psi}}{J(i\omega + K_{VM}/J_M)}\right\}^n}{1 - \left\{\frac{r e^{-i\psi}}{J(i\omega + K_{VM}/J_M)}\right\}}\right]$$

As $n \to \infty$ the series for $\dot{\theta}_n$ converges provided

$$\left|\left\{\frac{r e^{-i\psi}}{J(i\omega + K_{VM}/J_M)}\right\}\right| < 1$$

Thus

$$\underset{n\to\infty}{\dot{\theta}_n} = \mathcal{I}\left\{\frac{A e^{i\omega t}}{J(i\omega + K_{VM}/J_M)}\right\} \times \left[\frac{1}{1 - \frac{r e^{-i\psi}}{J(i\omega + K_{VM}/J_M)}}\right]$$

Putting

$$R = \frac{J_M}{J\sqrt{K_{VM}^2 + (\omega J_M)^2}}$$

and

$$\phi = \tan^{-1}\left(\frac{\omega J_M}{K_{VM}}\right)$$

and writing $\dot{\theta}$ for $\underset{n\to\infty}{\dot{\theta}_n}$

$$\dot{\theta} = T_q \frac{R e^{-i\phi}}{\{1 - rR e^{-i(\phi+\psi)}\}} \qquad (33)$$

In the absence of backlash the fundamental component of $\dot{\theta}$ becomes:

$$\dot{\theta} = T_q \frac{R e^{-i\phi}}{\{1 - r' R e^{-i\phi}\}}$$

Defining the describing function N_1 by

$$N_1 = \frac{\dot{\theta} \text{ when backlash is present}}{\dot{\theta} \text{ when backlash is absent}}$$

then

$$N_1 = \frac{\{1 - r' R e^{-i\phi}\}}{\{1 - rR e^{-i(\phi+\psi)}\}} \qquad (34)$$

Appendix II. Correction Factor for the Fundamental Component Describing Function Due to the Presence of Harmonics

If the torque developed by the motor is harmonic-free it follows from equation 31 that the harmonics of $J(D + K_{VM}/J_M)\dot{\theta}$ must vanish identically with the harmonics of $(1-\alpha/\alpha)K_{VL}\dot{\theta}_L$.

Thus $\dot{\theta}_L$ and $\dot{\theta}$ are defined by:

$$\dot{\theta} = \sum_{n=1}^{\infty} a_n \left\{e^{i(n\omega t + \phi_n)} + e^{-i(n\omega t + \phi_n)}\right\}$$

and

$$\dot{\theta}_L = \sum_{n=1}^{\infty} b_n \left\{e^{i(n\omega t + \beta_n)} + e^{-i(n\omega t + \beta_n)}\right\}$$

Equating harmonics of equation 31 to zero,

$$J(D + K_{VM}/J_M)\sum_{n=3}^{\infty} a_n \left\{e^{i(n\omega t + \phi_n)} + e^{-i(n\omega t + \phi_n)}\right\} = \frac{1-\alpha}{\alpha}(K_{VL})\sum_{n=3}^{\infty} b_n \left\{e^{i(n\omega t + \beta_n)} + e^{-i(n\omega t + \beta_n)}\right\}$$

Hence

$$2R_e\{a_n e^{i(n\omega t + \phi_n)}\} = 2R_e\left\{\frac{\frac{1-\alpha}{\alpha} K_{VL} b_n e^{i(n\omega t + \beta_n)}}{J(i\omega n + K_{VM}/J_M)}\right\}$$

From which is obtained

$$\phi_n = \beta_n - \tan^{-1}\left(\frac{n\omega J_M}{K_{VM}}\right)$$

and

$$a_n = \frac{b_n \frac{1-\alpha}{\alpha} \frac{K_{VL}}{n\omega J_M}}{\sqrt{1 + (K_{VM}/n\omega J_M)^2}}$$

Writing

$$\psi_n = \tan^{-1}\left(\frac{n\omega J_M}{K_{VM}}\right)$$

$$r_n = \frac{1-\alpha}{\alpha} K_{VL} b_n$$

and

$$R_n = \frac{1}{n\omega J\sqrt{1 + (K_{VM}/n\omega J_M)^2}}$$

gives

$$\dot{\theta}_n = r_n R_n e^{-i\psi_n}\dot{\theta}_{Ln} \qquad (35)$$

where θ_n is the nth harmonic of θ and θ_{Ln} is the nth harmonic of θ_L.

The harmonics of θ_L are determined by the nonlinear relationship between θ and θ_L. A first approximation to the harmonic content of θ_L can be calculated, using Fourier analysis, by assuming θ to be sinusoidal.

If θ_{Ln}' is the first approximation to the nth harmonic of θ_L then

$$\theta_{Ln}' = \frac{\cos\phi_1}{\pi}\left\{\frac{\sin(n+1)\theta_0}{n+1} - \frac{\sin(n-1)\theta_0}{n-1}\right\}\sin(n\omega t) + \frac{a_1\cos\phi_1}{\pi} - \left\{\frac{2}{n^2-1} - \frac{\cos(n-1)\theta_0}{n-1} + \frac{\cos(n+1)\theta_0}{n+1}\right\}\cos(n\omega t)$$

which may be written

$$\theta_{Ln}' = \mu_n e^{j\gamma_n}\theta_1 \tag{36}$$

in particular

$$\theta_{L1}' = \mu_1 e^{j\gamma_1}\theta_1$$

where

$$\mu_1 = \sqrt{a^2+b^2},\quad \gamma_1 = \tan^{-1}(-b/a)$$

Substituting from equation 36 into equation 35 for θ_{Ln}', a first approximation to the harmonics of θ is obtained.

Thus:

$$\theta_n' = r_n R_n e^{-j\psi_n}\mu_n e^{j\gamma_n}\theta_1 \tag{37}$$

Using Fourier analysis it can be shown that the effect of the nth harmonic of θ on the mth harmonic of θ_L is to produce extra mth harmonic of θ_L given by equation 38.

Extra mth harmonic of θ_L

$$= \frac{\sin(m\omega t)}{\pi}\left[a_n\left\{\frac{\sin(n+m)\theta_0}{n+m} - \frac{\sin(n-m)\theta_0}{n-m}\right\} + b_n\left\{\frac{2}{n^2-1} - \frac{\cos(n-m)\theta_0}{n-m} + \frac{\cos(n+m)\theta_0}{n+m}\right\}\right] + \frac{\cos(m\omega t)}{\pi}\left[a_n\left\{\frac{\cos(n+m)\theta_0}{n+m} + \frac{\cos(n-m)\theta_0}{n-m}\right\} - b_n\left\{\frac{\sin(n+m)\theta_0}{n+m} + \frac{\sin(n-m)\theta_0}{n-m}\right\}\right]$$

provided $n \neq m$. (38)

For $n=m$ the following is obtained:

$$\frac{1}{\pi}\left\{a_n\left(\pi-\theta_0+\frac{\sin 2n\theta_0}{2n}\right) - b_n\frac{\sin^2 n\theta_0}{n}\right\}\times \sin(n\omega t) + \frac{1}{\pi}\left\{-a_n\frac{\sin^2 n\theta_0}{n} + b_n\left(\pi-\theta_0+\frac{\cos 2n\theta_0}{2}\right)\right\}\cos n\omega t$$

Since a_n varies approximately as $1/n$ with the order of the harmonic it follows from equation 38 that for $n>m$ the contribution to the mth harmonic of θ_L is small compared with the mth harmonic when $n=m$.

It is thus assumed that the effect on the mth harmonic of θ_L is due entirely to the harmonics of order less than, or equal to, m. Thus in considering the third harmonic of θ_L only the fundamental and third harmonic of θ are considered to be effective.

Thus the extra harmonic in θ_L due to the harmonics of θ may be written as $\rho e^{i\xi}\theta_n$ where

$$\rho = \left[\left\{a_n\left(\pi-\theta_0+\frac{\sin 2n\theta_0}{2n}\right) - b_n\frac{\sin^2 n\theta_0}{n}\right\}^2 + \left\{b_n\left(\pi-\theta_0+\frac{\cos 2n\theta_0}{2n}\right) - a_n\frac{\sin^2 n\theta_0}{n}\right\}^2\right]^{1/2}$$

and

$$\xi = \tan^{-1}\left[\frac{\left\{b_n\left(\pi-\theta_0+\frac{\cos 2n\theta_0}{2n}\right) - a_n\frac{\sin^2 n\theta_0}{n}\right\}}{\left\{a_n\left(\pi-\theta_0+\frac{\sin 2n\theta_0}{2n}\right) - b_n\frac{\sin^2 n\theta_0}{n}\right\}}\right]$$

Hence a second approximation to the harmonics of θ_L is given as

$$\theta_{Ln}^2 = \theta_{Ln}' + \rho_n e^{i\xi}\theta_n'$$
$$= \theta_{Ln}'\{1+\rho R_n r_n e^{-j(\psi_n-\mathcal{D})}\}$$

From which

$$\theta_n^2 = r_n R_n e^{-j\psi_n}\theta_{Ln}^2$$
$$= r_n R_n e^{-j\psi_n}\theta_{Ln}'\{1+\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}$$

Using θ_n^2 the third approximation to θ_{Ln} may be calculated giving:

$$\theta_{Ln}^3 = \theta_{Ln}'\{1+\rho r_n R_n e^{-j(\psi_n-\mathcal{D})} + (\rho r_n R_n e^{-j(\psi_n-\mathcal{D})})^2\}$$

Continuing in this way one obtains the kth approximation as:

$$\frac{\theta_{Ln}^k}{\theta_{Ln}'} = [1+\rho r_n R_n e^{-j(\psi_n-\mathcal{D})} + \{\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}^2 + \ldots\{\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}^{k-1}]$$

Hence

$$\theta_{Ln}^k = \theta_{Ln}'\left[\frac{1-\{\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}^k}{1-\{\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}}\right] \tag{39}$$

Since $|rR|<1$ and $|r_n R_n|<1$, $\rho<1$, putting $\theta_{Ln}^k = \theta_{Ln}$ gives $k\to\infty$

$$\theta_{Ln} = \theta_{Ln}'\frac{1}{\{1-\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}} \tag{40}$$

Substituting for θ_{Ln}' from equation 36.

$$\theta_{Ln} = \frac{\mu_n e^{j\gamma_n}\theta_1}{\{1-\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}} \tag{41}$$

and from 35.

$$\theta_n = \frac{\mu_n e^{j\gamma_n} r_n R_n e^{-j\psi_n}\theta_1}{\{1-\rho r_n R_n e^{-j(\psi_n-\mathcal{D})}\}} \tag{42}$$

The effect of this harmonic on the fundamental component of θ_L is obtained from equation 38 by putting $m=1$.

Correction Factor Due to Third Harmonic

In this case the correction factor becomes

$$\frac{1}{\pi}\left[a_3\left\{\frac{\sin 4\theta_0}{4} - \frac{\sin 2\theta_0}{2}\right\} + b_3\left\{\frac{1}{4} - \frac{\cos 2\theta_0}{2} + \frac{\cos 4\theta_0}{4}\right\}\right]\sin\omega t + \frac{1}{\pi}\left[a_3\left\{\frac{\cos 4\theta_0}{4} + \frac{\cos 2\theta_0}{2}\right\} - b_3\left\{\frac{\sin 4\theta_0}{4} + \frac{\sin 2\theta_0}{2}\right\}\right]\cos\omega t$$

where a_3 and b_3 are given by:

$$a_3 + ib_3 = \frac{r_3 R_3 e^{-i\psi_3}\mu_1 e^{j\gamma_1}|\theta_1|}{\{1-\rho r_3 R_3 e^{-i(\psi_3-\mathcal{D})}\}}$$

Appendix III. Conditions for the Existence of a Limit Cycle in a Second-Order System with Output Velocity Feedback

When the input signal is zero, the equation of motion becomes:

$$J\ddot{\theta} + K_{VM}\dot{\theta}_M + K_{VL}\dot{\theta}_L + K_M\theta_M + K_L\theta_L = 0 \tag{43}$$

Writing $D = d/dt$,

$$\alpha = \frac{K_{VL}/J_L}{K_{VM}/J_M},\quad \gamma = \frac{K_L/J_L}{K_M/J_M}$$ in equation 43,

$$J\left\{D^2 + \frac{K_{VM}}{J_M}D + \frac{K_M}{J_M}\right\}\theta - \left\{\frac{1-\alpha}{\alpha}(K_{VL})D + \frac{1-\gamma}{\gamma}K_L\right\}\theta_L = 0 \tag{44}$$

It is wished to find a solution of equation 44 for which θ is periodic.

Assuming θ is periodic, then

$$\theta = \mathcal{S}A_s e^{j(s\omega t+\phi)}$$

Since the input is zero ϕ may be given any arbitrary value without loss of generality. Thus

$$\theta = \mathcal{S}A_s e^{js\omega t} \tag{45}$$

Further, when θ is periodic,

$$\theta_L = (a-ib)\theta = \sqrt{a^2+b^2}\,e^{-i\phi}\theta \tag{46}$$

Substituting from equations 45 and 46 into equation 44,

$$J\left\{D^2 + \frac{K_{VM}}{J_M}D + \frac{K_M}{J_M}\right\}\mathcal{S}A_s e^{js\omega t} - \left\{\frac{1-\alpha}{\alpha}(K_{VL})D + \frac{1-\gamma}{\gamma}K_L\right\}\times (a-ib)\mathcal{S}A_s e^{js\omega t} = 0$$

$$\mathcal{S}\left\{-\omega^2 + \frac{K_{VM}}{J_M}i\omega + \frac{K_M}{J_M}\right\}A_s e^{js\omega t} - (a-ib)\mathcal{S}A_s e^{js\omega t}\times \left\{\frac{1-\alpha}{\alpha}(K_{VL})i\omega + \frac{1-\gamma}{\gamma}K_L\right\} = 0$$

$$\left\{-\omega^2 + \frac{K_{VM}}{J_M}i\omega + \frac{K_M}{J_M}\right\} - (a-ib)\times \left\{\frac{1-\alpha}{\alpha}(K_{VL})i\omega + \frac{1-\gamma}{\gamma}K_L\right\} = 0$$

Equating real parts:

$$\left\{(K_M/J_M-\omega^2)J-\frac{1-\gamma}{\gamma}(a)K_L-\frac{1-\alpha}{\alpha}(K_{VL})\,b\omega\right\}=0 \quad (47)$$

Equating imaginary parts:

$$\frac{J}{J_M}\omega K_{VM}-\frac{1-\alpha}{\alpha}(K_{VL})\omega a+\frac{1-\gamma}{\gamma}(K_L)b=0 \quad (48)$$

From which is obtained:

$$\omega=\frac{\frac{1-\gamma}{\gamma}(K_L)b}{\left\{\frac{1-\alpha}{\alpha}(K_{VL})a-\frac{J}{J_M}K_{VM}\right\}} \quad (49)$$

$$b^2=\frac{\left\{\frac{K_M}{J_M}J-\frac{1-\gamma}{\gamma}(a)K_L\right\}\times\left\{\frac{1-\alpha}{\alpha}(K_{VL})(a)-\frac{J}{J_M}K_{VM}\right\}^2}{\left[J\left(\frac{1-\gamma}{\gamma}(K_L)\right)^2+\frac{1-\alpha}{\alpha}\left(\frac{1-\gamma}{\gamma}\right)\times(K_L)(K_{VL})\left\{\frac{1-\alpha}{\alpha}(K_{VL})a-\frac{J}{J_M}K_{VM}\right\}\right]} \quad (50)$$

Also, a and b are related through the nonlinearity such that:

$b=F_1(\theta_0)$, $a=F_2(\theta_0)$ (equations 9 and 10).

Hence

$$b=f(a) \quad (51)$$

A periodic solution of equation 44 exists provided equations 49, 50, and 51 can be satisfied simultaneously with ω positive.

Special Cases

All Reset from the Load. In this case

$\gamma=\infty$, $\alpha=\infty$

$K_{VL}=K_V$, $K_L=K$, and $K/J=\omega_0^2$

Substituting in equations 49 and 50 therefore

$$b^2=\frac{a^2(K_V/\omega_0 J)^2}{\left\{1+a(K_V/\omega_0 J)^2\right\}} \quad (52)$$

$$\frac{\omega}{\omega_0}=\left(\frac{b}{a}\right)\left(\frac{\omega_0 J}{K_V}\right) \quad (53)$$

It follows from equation 53 that ω/ω_0 is always positive, hence if equations 52 and 51 can be satisfied simultaneously a periodic solution exists. The amplitude and frequency of the periodic solution are clearly dependent on $K_V/\omega_0 J$.

It may be shown from the form of equation 51 that for all finite values of $K_V/\omega_0 J$ a periodic solution exists.

All Reset from the Motor. In this case

$K_L=0$, $K_{VL}=0$, $K_M=K$, $K_{VM}=K_V$

and

$$\frac{1-\gamma}{\gamma}(K_L)=\frac{J_L}{J_M}(K),\ \frac{1-\alpha}{\alpha}(K_{VL})=\frac{J_L}{J_M}$$

From which

$$b^2=\frac{\frac{J_L}{J_M}\left\{\frac{J_M}{J_L}+(1-a)\right\}\left\{a-\frac{J}{J_L}\right\}^2}{\left\{\left(\frac{\omega_0 J}{K_V}\right)^2+\frac{J_L}{J_M}\left(a-\frac{J}{J_L}\right)\right\}} \quad (54)$$

$$\frac{\omega}{\omega_0}=\frac{\omega_0 J}{K_V}\frac{b}{\{a-J/J_L\}} \quad (55)$$

Since $a\leq 1$ and $J/J_L\geq 1$ equation 55 cannot be satisfied for $\omega>0$. Thus in this case no periodic solution is possible.

Positional Reset from the Load, Velocity Signal from the Motor. In this instance one obtains

$K_{VM}=K_V$, $K_{VL}=0$, $K_L=K$, $K_M=0$

and

$$\frac{1-\alpha}{\alpha}K_{VL}=K_V(J_L/J_m)$$

Hence equations 49 and 50 become:

$$b^2=\frac{a(a-J/J_L)^2(K_V/\omega_0 J)^2}{\{(1-J/J_L)^2-(a-J/J_L)(K_V/\omega_0 J)^2\}} \quad (56)$$

$$\frac{\omega}{\omega_0}=\frac{\omega_0 J}{K_V}\frac{b}{J_L/J_M(J/J_L-a)} \quad (57)$$

From equation 57 it is seen that ω is positive for all a. Hence equations 56 and 51 determine whether a periodic solution exists.

In this case two periodic solutions are predicted for a given value of $K_V/\omega_0 J$. It has been shown that one of these solutions represents a divergent point of oscillation and that the value of $K_V/\omega_0 J$ for which $da/d(K_V/\omega_0 J)=\infty$ gives the minimum value of $K_V/\omega_0 J$ to prevent the existence of a periodic solution.

Appendix IV. Determination of Stability Limit

For any value of J/J_L equation 56 defines a curve in the $(a\sim K_V/\omega_0 J)$ plane.

The stability limit is that value of $K_V/\omega_0 J$ for which the tangent to these curves is vertical.

Hence $d(K_V/\omega_0 J)/da=0$ defines the stability limit. Differentiating equation 56 with respect to a and putting $d(K_V/\omega_0 J)/da=0$ leads to:

$$\frac{J}{J_L}=\frac{2a(a-bda/db)+\sqrt{b^2}da/db\{da/db(a^2-b^2)+2ab\}}{2\left(a-\frac{b}{2}da/db\right)} \quad (58)$$

Thus if a is specified, b and da/db are calculable from equation 51 and J/J_L can be determined.

Thus J/J_L may be found as a function of the value of a at which oscillations cease.

The smallest value of a arises when $J/J_L\to\infty$

In this case:

$$a=\frac{b}{2}\frac{da}{db} \quad (59)$$

from which $a=0.184$, $b=0.23$.

Substituting these values into equation 56 gives

$$\frac{K_V}{2\omega_0 J}=0.536$$

To determine the stability limit for any finite value of J/J_L, a is specified and J/J_L is calculated from equation 58. Substitution in equation 56 then gives $K_V/\omega_0 J$.

Appendix V. Nomenclature

J_M = moment of inertia on motor side of backlash

J_L = moment of inertia on load side of backlash

$J=J_M+J_L$

θ_M = motor position

θ_L = load position

θ = position of the center of mass

K_{VM} = viscous friction torque coefficient for the motor side of the backlash

K_{VL} = viscous friction torque coefficient for the load side of the backlash

$K_V=K_{VL}+K_{VM}$ = total viscous friction torque coefficient

T_e = torque developed by the motor

$\alpha=K_{VL}/J_L/K_{VM}/J_M$

a = amplitude of the fundamental component of θ_L in phase with the fundamental component of θ/amplitude of fundamental component of θ

b = amplitude of fundamental component of θ_L lagging the fundamental component of θ by $\pi/2$/amplitude of fundamental component of θ

$\psi=\tan^{-1}(b/a)$

$r=(1-\alpha/\alpha)(K_{VL})\sqrt{a^2+b^2}$

θ_0 = separation time/half-periodic time

$1/2\pi\sqrt{K_M/J_M}$ = the undamped natural frequency of the motor alone

$1/2\pi\sqrt{K_L/J_L}$ = the undamped natural frequency of the load when the motor has zero inertia

$\gamma=\frac{K_L/J_L}{K_M/J_M}$

$\zeta=K_V/2\omega_0 J$

$\omega_c/2\pi$ = the undamped natural frequency of the system without backlash

$r'=(1-\alpha/\alpha)(K_{VL})$

θ_{BL} = backlash angle

$R=J_M/J\sqrt{K_{VM}{}^2+(\omega J_M)^2}$

$\phi=\tan^{-1}(\omega J_M/K_{VM})$

References

1. The Effects of Backlash and of Speed-Dependent Friction on the Stability of Closed-Cycle Control Systems, A. Tustin. *Journal*, Institution of Electrical Engineers, London, England, pt. IIA, 1947, pp. 143–51.

2. The Dual-Input Describing Function and Its Use in the Analysis of Non-Linear Feedback Systems, J. C. West, J. L. Douce, R. K. Livesley. *Proceedings*, Institution of Electrical Engineers, 1956, pt. B.

3. Backlash and Resilience Within the Closed Loop of Automatic Control Systems, J. H. Liversidge. "Automatic and Manual Control" (book), Butterworth's Scientific Publications, London, England, 1952.

4. Backlash in Position Control Systems with Load Resonance. "Servo Mechanisms and Regulating System Design, Vol. II" (book), John Wiley & Sons, Inc., New York, N. Y.; Chapman & Hall, Ltd., London, England, 1955.

5. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, Ralph J. Kochenburger. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 270–84.

6. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements,

E. Calvin Johnson. *Ibid.*, vol. 71, pt. II, July 1952, pp. 109–81.

7. Classical Mechanics (book), D. E. Rutherford. Oliver & Boyd, Ltd., London, England, 1951.

8. Linearization of Non-Linear Elements for Small Departures. "Servo Mechanisms and Regulating System Design, Vol. II" (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc.; Chapman & Hall, Ltd.

9. Non-Linear Vibrations in Mechanical and Electrical Systems (book). J. J. Stoker. Interscience Publishers, London, England, 1950.

Discussion

S. C. Browning (Nelson Research Laboratories, The English Electric Company, Ltd., Stafford, England): The paper shows remarkably good agreement between the theory presented and the experimental results on a simulated second-order system.

From a qualitative viewpoint, divided reset can be seen to have a stabilizing effect, since it is intermediate between motor reset, in which case the backlash is outside the active loop, and load reset, in which case steady oscillations or limit cycles may occur.

It is considered that one of the principal advantages of the describing-function technique is its simplicity and that it allows the effect of altering various parameters to be easily visualized. The designer of control systems is generally interested in avoiding limit cycles rather than making an exact calculation of their frequency and amplitude. From this viewpoint the added complication of corrections to the describing function is not justified.

The author states that it was not found possible to evaluate the describing functions for combined nonlinear friction and backlash. A first study of this describing function has in fact been published by Satyendra.[1]

It is worth noting that divided or motor reset is only applicable where linear elements follow the backlash and friction.

Reference

1. Describing Functions Representing the Effect of Inertia, Backlash, and Coulomb Friction on the Stability of an Automatic Control System—1, K. N. Satyendra. *AIEE Transactions*, vol. 75, pt. II, Sept. 1956, pp. 243–49.

E. A. Freeman: The author agrees with Mr. Browning that the main advantage of the describing function technique is the simplicity with which one may predict possible sustained oscillations. However, it must be borne in mind that the designer of practical control systems must have an estimate of the accuracy of his predictions so that he may design into the system the facility to vary certain system parameters over certain ranges. This facility is particularly important when the system specification requires possible limit cycles to have amplitudes less than a specified value. The correction factor developed by the author allows the design engineer to estimate the accuracy of his predictions and if necessary use the more accurate describing functions to improve his predictions. In this way it is suggested that systems designed to be stable will not erroneously exhibit sustained oscillations of amplitude greater than the maximum permitted.

The paper presented by Dr. Satyendra[1] makes two simplifying assumptions which restrict the consequent application of the work presented.

First, he assumes that the load only is subject to coulomb friction. In many practical control systems with backlash the driving member, because of its higher speed, experiences comparatively larger frictional forces than the load.

Second, he assumes that the driving member mass is infinitely large and consequently the position of the driving member is identical to the position of the center of mass. All practical control systems, however, have finite driving member mass and, therefore, the motion of the load affects that of the driver. Because of this the position of the center of mass is an unknown. The theory presented by the present author indicates that with finite masses and linear speed-dependent friction the position of the center of mass may change appreciably from that assumed by Dr. Satyendra. Further, it has been illustrated that these changes can, in sensibly practical systems, lead to very misleading results if an approach similar to that of Dr. Satyendra is employed.

A method of analysis which overcomes these difficulties with nonlinear friction was not found possible initially. Recently the author has developed the method presented in the paper to give an approximate account of the effects of backlash and nonlinear friction.

Reference

1. See reference 1 of the Browning discussion.

LATE DISCUSSIONS

Discussion of Characteristics of an Electric Resistance Furnace Load, W. A. Stelzer, pages 570-72

A. M. Killin (Electro Metallurgical Company, Niagara Falls, N. Y.): Mr. Stelzer in his paper discusses the fundamentals of a graphitizing furnace and its electric resistance load. The furnace that he had in mind is 355 inches long, 87 inches wide, and 85 inches high and operates at a maximum load of 2,700 kw.

The first step in the graphitizing process is called "baking" and requires an electrical source of approximately 47 volts with a gradual increased current from about 1,100 to 1,300 amperes over a period of 408 hours to a current from 6,000 to 6,500 amperes.

In the second step of the graphitizing process voltage across the primary terminals is increased from 65 to 70 volts with a furnace current of about 17,000 amperes for a period of around 24 hours. The voltage is then increased to 75–90 volts with a current of 35,000 to 40,000 amperes for a period of about 40 hours.

Current for the baking process is provided from a 400-kva single-phase 60-cycle transformer. The conductors between the transformer and furnace are interlaced copper bars of 1/2-inch by 6-inch cross section provided with suitable switches at the end of the furance so that the baking current can be disconnected. After the baking process the graphitizing transformer of 3,400-kva capacity is connected to the furnace through water-cooled cables. This transformer has a novel arrangement in that it and series capacitor transformers are mounted on a standard-gage specially designed railway truck with motor drive. This permits the movement of the transformer assembly to each of the 16 graphitizing furnaces and permits a decrease investment in copper bus bars.

Mr. Stelzer points out the power-factor difficulties in a graphitizing furnace. The power factor varies approximately from 82% at the start of the graphitizing run and gradually falls off to around 55% at the end of the run due to the decrease in the resistance of the furnace. Through the use of series capacitor transformer and two capacitors power-factor correction is made.

H. Tauber (The Detroit Edison Company, Detroit, Mich.): The author has described a resistance-type furnace load and the manner in which he ingeniously overcomes some equipment limitations. The use of series capacitors for power-factor improvement, and the interchange of two transformers for the complete process, are perfect examples of utilizing available equipment to their utmost.

The power supply designer will be most interested to note that the load characteristics of such furnaces are dependent on the materials used to pack the furnace, and their proportions. Here is a load that varies between 50 kva and 3,600 kva at power factors between 97% and 40% during the course of a complete baking and graphitizing period, and might possibly cause voltage regulation problems, depending on the nature of the power supply.

I note that this load is served from a 13.8-kv single-phase bus, presumably connected line-to-line. Since some 13-kv systems are of the 4-wire type, I would like to point out that, for a given kva single-phase load, the voltage drop is considerably less when the connection is line-to-line than if it is line-to-neutral. In cases where regulation or flicker may be a problem, the line-to-line connection becomes attractive even though higher voltage rated equipment must be used.

The author deserves the thanks of power supply designers for describing the load characteristics of the graphitizing furnace and for the explanation of some interesting methods of utilizing available equipment.

W. A. Stelzer: I thank Mr. Killin and Mr. Tauber for their discussions. It should be noted that all changes of currents and voltages in the baking and graphitizing process described are slow changes over the period of the run. When voltage and current adjustments are made by means of the manual underload tap-changer controls, there are disturbances on the 13.8-kv supply line. No complaints have been made because of these disturbances and we therefore conclude they are not objectionable to the other users of power on the line.

The process load is served from the 13.8-kv line, phase-to-phase or line-to-line as Mr. Tauber has surmised, and regulation or flicker has not become a problem.

Discussion of Differential Analyzer Aids Design of Electrical Utility Automatic Dispatching System, L. K. Kirchmayer, pages 572–79

Arthur Klopfenstein (Southern California Edison Company, Los Angeles, Calif.): This is a very interesting and informative paper and is a significant addition to the fine series of papers on power system generation control of which Dr. Kirchmayer has been an author.

The automatic dispatching system in use to control most of the generation of the Southern California Edison Company provides no feedback loop from either the control valves or from the speed changer to the unit controller as is indicated on Fig. 14 of this paper. No additional feedback loops are needed on our hydroturbine governors. These governors have a dead band that is only about one third of the AIEE–American Society of Mechanical Engineers standard dead band. These hydro governors do have a mechanical feedback loop from gate to input pilot valve which contain dashpots that cause a temporary increase in governor droop during very rapid changes. The time constant for our hydrogeneration changes are determined primarily by penstock time constant instead of governor delay.

We have found that a time delay of from 30 to 60 seconds is required to overcome the backlash in two of our 150-megawatt General Electric steam-turbine governors when initiating a load change at 4 megawatts per minute with our automatic dispatching system. These large delays might be reduced by adding control valve to speed-changer feedback loops as indicated in Fig. 14. This backlash which causes control delay is also the cause of a large governor dead band that is seldom exceeded by deviations of our system frequency. To obtain frequency changes large enough to measure the extent of this dead band, a large block of generation was suddenly dropped and picked up on a hydro unit to cause frequency changes that exceeded dead band.

When nearly instantaneous, 10% of rating changes were made in steam-turbine output with the speed-changer handwheel, the backlash was quickly overcome and time constants of governor steam flow and generation changes were measured that are in the same order of magnitude as time constants used by Dr. Kirchmayer in his differential analyzer study.

Methods and quantitative results from transient and frequency response tests on Southern California Edison Company hydroelectric and steam-turbine generating units will be presented in technical papers in the near future.

L. K. Kirchmayer: I wish to thank Mr. Klopfenstein for his comments. The industry will indeed find of great interest Mr Klopfenstein's future paper on methods and response tests on hydro- and steam-turbine generating units.

G. Clark of the General Electric large steam turbine department and I have recently reviewed some of the Southern California Edison tests. Mr. Klopfenstein is correct in indicating that two of their General Electric units have dead bands in excess of established standards. We believe that the majority of the dead band is due to friction in the control-valve operating mechanism. We expect that the pressure-stroke characteristic of the servo pilot valve can be improved to better combat the friction and to bring the dead band below the value of 0.06% speed change allowed by the AIEE Standard.[1]

The time delay associated with the response of the turbine governor is directly related to the governor dead band and inversely proportional to rate of change of the control signal. Assuming the allowable governor dead band of 0.06% and a control rate of 4 megawatts power change per minute, approximately a 30-second time delay would be incurred. With a control rate of 2 megawatts per minute, the time delay would be approximately 60 seconds. The delays discussed in the foregoing are inherent with the control signal arrangement described by Mr. Klopfenstein for turbines built according to standard steam turbine practices.

There are methods by which the turbine control system can be improved to shorten the delay time; several of these methods would require special developments to solve several problems, one of which concerns control stability. One of the direct ways of reducing delays is through the use of other control arrangements which incorporate feedback from the speed level changer and the turbine valve position or generator power output.

Reference

1. Specification for Speed Governing of Steam Turbines for Generators (500 Kw and Up). *AIEE Standard no. 600*, May 1949.

Discussion of Motor Field Control of Large D-C Reversing Mill Motors, L. R. Hulls and G. H. Samuel, pages 579–85

H. N. Snively (General Electric Company, Schenectady, N. Y.): The authors are to be congratulated on their analysis of motor field control of large d-c reversing mill motors.

To get optimum performance from such motors it is most essential that the proper relationship between motor field current and motor armature current be maintained as described in the paper.

From an application standpoint, the control of armature voltage is just as important as the control of field flux. In our own analytical studies, although our approach to field control was somewhat different, we have considered the entire problem, including nonlinearities. Thus, we could study on the computer actual motor reversals from top-to-top speed. The circuits so developed have now been in service for several years, and high-speed oscillograph recordings of transient voltages and currents compare exactly with those predicted by our analog computer.

L. R. Hulls and **G. H. Samuel:** We agree with Mr. Snively that armature voltage control is just as important as the control of motor field flux. The problems involved, however, are much easier to solve and are well understood. In our paper we attempt to shed some light on the more complex motor field control.

The study could have been made to include armature voltage control by substituting the output of a voltage regulator analog for the fixed voltage obtained from *R3* in Fig. 7; but, for the sake of simplicity, we generally prefer to carry out separate studies for the generator voltage regulator and the motor-field-current regulator.

We would like to thank Mr. Snively for his remarks on our paper.

(Additional Discussion will appear in the May 1959 issue.)